

The London School of Economics and Political Science

Contractarianism's Dilemma

*— On the normativity of contemporary
contractarian theories*

Baldwin Bon-Wah Wong

A thesis submitted to the Department of Government of
the London School of Economics and Political Science
for the degree of PhD in Government, London,
December 2011

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

The copyright of this thesis rests with the author. Quotation from it is permitted, provided that full acknowledgement is made. This thesis may not be reproduced without my prior written consent.

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of **94359** words.

Abstract

Contractarianism has a distinguished history and is one of the most influential schools of thought nowadays, yet there are only few general discussions about this school. The research question which intrigues me is whether contemporary contractarianism can provide a satisfactory normative justification for political principles. I argue that contractarianism, as a methodology, consists of three elements: a conception of practical reason, hypothetical contractors, and a hypothetical contract. Based on various conceptions of practical reason, different contractarian models can be developed. In this thesis, I examine three possible contractarian models: Hobbesian contractarianism (represented by David Gauthier), Kantian contractarianism (represented by T. M. Scanlon) and hybrid contractarianism (represented by John Rawls). I diagnose the shortcomings of these three existing models respectively. Hobbesian contractarianism assumes a conception of rationality, hence it conceives hypothetical contractors as individual utility-maximizers, and the hypothetical contract as a mutually advantageous agreement. Kantian contractarianism assumes a conception of reasonableness, hence it conceives hypothetical contractors as moral persons who would behave in a way which could be justifiable to one another, and the hypothetical contract as an agreement that no one could reasonably reject. These two models fail since their conceptions of practical reason are too one-sided: the former overlooks reasonableness, whereas the latter overlooks rationality. Due to their one-sidedness, these models can at best justify political principles that are general but not overriding. Hybrid contractarianism avoids this problem by assuming that hypothetical contractors were both rational and reasonable and proving that rationality and reasonableness would justify the same hypothetical contract. However, in order to show the congruence between rationality and reasonableness, this model inevitably assume substantial, controversial conceptions of practical reason. Hence, hybrid contractarianism can at best justify political principles that are overriding but not general. The failures of these three models show the limit of this methodology. No matter how contractarians construct

their models, their models are subject to the fatal dilemma of choosing between generality and priority. While these two properties are necessary for political principles, this implies that contractarianism does not have the resources to offer a satisfactory normative justification for political principles.

Acknowledgement

This thesis is meant to be a critical exposition of and commentary on contractarianism, which is one of the most influential schools of thought in contemporary political philosophy. I discuss three most eminent members of this school in the twentieth century—John Rawls, T. M. Scanlon and David Gauthier. In fact, this school of thought drew my interest since I first got in touch with political philosophy six years ago. In my intellectual journey, contemporary contractarianism always occupies an irreplaceable role. I began to be interested in political philosophy in my third undergraduate year, and Rawls' *A Theory of Justice* was my first philosophical book. My master thesis was about Rawls' *Political Liberalism*. After I came to LSE for my PhD, I broadened my study to include other contractarians. I was obsessive with the history of social contract in my first year. I then devoted my second year to studying Gauthier and my third year to Scanlon. Finally most of my thoughts were written down in this thesis during my fourth year. Hence, this thesis is like a crystal of my intellectual journey in the last few years. My feelings are complex when I reach the end of this journey, but the strongest feeling should be a sense of satisfaction. Although this thesis is no doubt imperfect, I am greatly satisfied with it and take it to be a summary of my philosophical reflections in these six years.

However, it is unfair to say that this thesis is the product of my own effort. I always believe that interaction with others is necessary for philosophical reflection. Hence I would like to take this opportunity of offering my sincere gratitude to people who have helped me so much at every stage of the process. I am benefited immeasurably from the guidance of Chandran Kukathas, who supervised this thesis. Throughout my PhD year, I have received extraordinary supports from him in so many ways. I am greatly benefited from his detailed and thoughtful feedbacks which always push me to articulate my ideas with the greatest clarity. By discussing with him, I learned how to analyze philosophical issues in a very serious and rigorous way. His kindness also touches me because, every time when I felt depressive and doubted the values of my

works he was the first one who came out to encourage and assist me. I am also grateful to Paul Kelly, who is my advisor. My trichotomy of contractarian theories is highly indebted to my exchanges with him. I also learnt a great deal from Zsuzsanna Chappell, Philip Cook, Katrin Flikschuh, James Gledhill, Alex Leveringhaus, Christian List, Anne Philips and Kai Spiekerman. Thanks LSE for providing such an intellectual stimulating environment for me to engage in my research.

I have been fortunate to work with a brilliant group of fellow doctoral students. Many friends have offered invaluable advices at various stages. I thank all of them, including, Aslan Amani, Brian Chen, Katy Dineen, Muriel Kahane, Andrej Keba, Camillia Kong, Edward Hall, Pietro Maffetone, Alison Mallard, Yonathan Reshef, Esha Senchaudhuri, Rachel Tsang, Luke Ulas, Jeremy Williams and James Wong. I would like to single out Edward Hall and Luke Ulas in particular, who kindly proofread some of my chapters and offer a lot of insightful comments. Also, I owe a great debt to Pietro Maffetone and Yonathan Reshef, who tolerate countless discussions about some topics with which the book engages. Indeed, without Yon, Chapter 7 would be highly different from what it is now.

I am grateful to all those who gave comments when some chapters were presented earlier at a number of conferences and workshops. Especially thanks to the stimulating discussions with Ian Carter, David Gauthier, Gerald Gaus, Stephen Macedo, Matt Matravers, Christopher Morris, Martin O'Neill, Geoffrey Sayre-McCord, Hillel Steiner, Adam Swift, Laura Vallentini and Burno Verbeek.

I also owe a great intellectual deal to Po-Chung Chow and Thomas Cham, who inspired my interest in political philosophy during my undergraduate age. My journey in political philosophy began from day-and-night discussions with them. Had I not encountered them I would have a totally different life.

I am also thankful to Adrian Chiu, my best friend who has offered me enormous assistance since I first came to England. Thanks for his unconditional support and care during the difficult time of writing this thesis.

Finally, I should like to thank my parents, Lawrence Wong and Amy Wong. Throughout these years they offered more love and backing than I could ever hope for. I am sorry that I have been overseas for most of the time in these five years because of this degree. Despite this, they are still willing to respect my choice and support me to choose this 'high-risk-low-return' path. I deeply appreciate to their unfailing love.

Table of Contents

Chapter 1 Introduction	15
1.1 A very brief history of contractarianism	18
1.2 Various functions of contractarianism.....	23
1.3 The nature of political principle	28
1.4 Method and outline.....	311
1.5 Contractarianism: a patriarchal theory?.....	36
Chapter 2 Why Contractarianism?	41
2.1 Introduction	41
2.2 Three virtues of contractarianism.....	44
2.2.1 The virtue of naturalism	44
2.2.2 The virtue of public justifiability	48
2.2.3 The virtue of respecting plurality	50
2.3 The ‘standard critique’: why am I bound by a hypothetical contract?	52
2.3.1 The necessity of idealisation	53
2.3.2 The standard critique.....	54
2.3.3 The orthodox reply and its inadequacy	56
2.4 The normativity of hypothetical contract	59
2.4.1 The normativity of practical reason	60
2.4.2 Hypothetical contract and practical reason	63
2.4.3 The problem of the standard critique	66
2.5 Two conditions: the condition of generality and the condition of priority.....	70
2.5.1 The condition of generality	71
2.5.2 The condition of priority	73
2.5.3 Case study: John Harsanyi	77
2.6 Conclusion.....	80
Chapter 3 Three Models of Contractarianism	83
3.1 Introduction	83
3.2 Hobbesian contractarianism and the conception of rationality	86
3.2.1 The conception of rationality	86
3.2.2 Hobbesian contractarianism.....	90

Chapter 1 Introduction

3.3	Kantian contractarianism and the conception of reasonableness	92
3.3.1	The conception of reasonableness.....	93
3.3.2	Kantian contractarianism	96
3.4	The myth of dichotomy	98
3.4.1	A re-interpretation of Rawls' contractarianism	100
3.4.2	The conception of reasonableness in Rawls' contractarianism	101
3.4.3	The conception of rationality in Rawls' contractarianism.....	103
3.4.4	Between Hobbesian and Kantian contractarianism	106
3.5	The third possibility: hybrid contractarianism and the dual conception of practical reason	109
3.5.1	The dual conception of practical reason	110
3.5.2	Hybrid contractarianism.....	114
3.6	Conclusion.....	116
Chapter 4 Hobbesian contractarianism: Gauthier		121
4.1	Introduction	121
4.2	Hobbesian contractarianism: an elaboration	124
4.2.1	Characteristics of hypothetical contractors	123
4.2.2	The state of nature: the Prisoners' Dilemma.....	127
4.2.3	The hypothetical contract: the Lockean proviso and the principle of minimax relative concession	130
4.3	Hobbesian contractarianism and the condition of generality	133
4.3.1	The general acceptability of the conception of rationality.....	134
4.3.2	Presupposing hedonism?.....	137
4.3.3	Presupposing egoism?.....	139
4.3.4	Too unrealistic?.....	140
4.4	Reasonable rejectability of Hobbesian contractarianism	144
4.4.1	The reasonable aspect of practical reason.....	145
4.4.2	Why Hobbesian contractarianism is to be reasonably rejected? (1): the powerlessness critique.....	146
4.4.3	Why Hobbesian contractarianism is to be reasonably rejected? (2): the inadequacy of the Lockean proviso.....	150
4.5	Can reasonableness be reduced to a preference?.....	155
4.5.1	The reductionist argument	156
4.5.2	The other-regarding feature of reasonableness	158

Chapter 1 Introduction

4.5.3	The cost of broadening the definition of individual utility	160
4.6	Conclusion.....	163
Chapter 5 Kantian Contractarianism: Scanlon and Barry.....		167
5.1	Introduction	167
5.1.1	A brief comparison between Scanlon and Barry	170
5.2	Kantian contractarianism: an elaboration.....	172
5.2.1	Characteristics of hypothetical contractors	172
5.2.2	The state of nature: the reason-exchanging conversation	176
5.2.3	The hypothetical contract: principles which no one could reasonably reject	178
5.3	Kantian contractarianism and the condition of generality.....	181
5.3.1	The conception of reasonableness and the sense of guilt.....	181
5.3.2	Is the conception of reasonableness neutral?	184
5.3.3	The empirical evidence of general acceptability.....	186
5.4	Rational rejectability of Kantian contractarianism.....	187
5.4.1	The conception of rationality in Kantian contractarianism.....	187
5.4.2	Why Kantian contractarianism is to be rationally rejected? (1): the ineffectiveness critique	190
5.4.3	Why Kantian contractarianism is to be rationally rejected? (2): the self-contradiction critique	192
5.5	Should reasonableness be considered prior? (1): two arguments.....	194
5.5.1	The leave-room argument	196
5.5.2	The proper-understanding argument.....	198
5.6	Should reasonableness be considered prior? (2): the great-value argument.....	200
5.6.1	The great-value argument	201
5.6.2	Valuableness and priority.....	202
5.6.3	The externalist account and its costs.....	204
5.7	Should reasonableness be considered prior? (3): the scepticism argument.....	208
5.8	Conclusion.....	211
Chapter 6 Hybrid Contractarianism (1): the earlier Rawls.....		215
6.1	Introduction	215
6.2	Hybrid contractarianism: an elaboration	217

Chapter 1 Introduction

6.2.1	The problem of the orthodox interpretation	218
6.2.2	A re-interpretation of Rawlsian hypothetical contractors: free and equal persons	219
6.2.3	The hypothetical contract: justice as fairness	228
6.3	Hybrid contractarianism and the condition of priority	234
6.3.1	The priority of justice as fairness	234
6.3.2	The conflict between two highest-order interests	238
6.3.3	The maximizing nature of the highest-order interest	241
6.4	The cost of satisfying the condition of priority	244
6.4.1	The problem of generality	245
6.4.2	Defence (1): Kantian interpretation as a fact	248
6.4.3	Defence (2): the necessity of two moral capacities.....	250
6.5	Conclusion.....	254
Chapter 7 Hybrid Contractarianism (2): the later Rawls.....		257
7.1	Introduction	257
7.2	The revision of hybrid contractarianism: overlapping consensus	259
7.2.1	The change of hypothetical contractors: scope and motivation.....	259
7.2.2	Hybrid contractarianism and the condition of priority.....	264
7.2.3	Hybrid contractarianism and the condition of generality.....	267
7.3	The problem of generality revisited	270
7.3.1	The limit of scope: from the world to liberal democratic societies	271
7.3.2	The limit of scope: from real liberal democratic societies to ideal well-order society.....	273
7.3.3	Weakly just citizens and strongly just citizens	276
7.3.4	The tension between freestandingness and overridingness.....	283
7.4	Creating ideal citizens by education.....	291
7.4.1	Public political culture and background culture	291
7.4.2	Examining the influence of families and associations	293
7.5	The desirability of justice as fairness	298
7.5.1	The collective-social-good argument.....	299
7.5.2	The self-respect argument.....	301
7.6	Conclusion.....	305
Chapter 8 Conclusion.....		309
8.1	The dilemma of contractarianism.....	309

Chapter 1 Introduction

8.2	Contributions to the existing approaches	313
8.3	The future of contractarianism?.....	316
Bibliography		321

List of Tables

Table 1	Rationality and Reasonableness.....	99
Table 2	Three Models of Contractarianism.....	116
Table 3	Prisoners' Dilemma.....	129
Table 4	A version of Allais Paradox.....	142
Table 5	Contractarianism's Dilemma.....	310

Chapter 1 Introduction

The justification of political principles is a long-standing question in political philosophy.¹ Political philosophers have used different methodologies to justify that certain principles have normativity, that is, people are bound by these principles and *ought* to follow them. Contractarianism is one of the most commonly used methodologies. I take contractarianism to be a methodology of justifying political principles by arguing that these principles represent an (appropriate) agreement between individuals who are ruled by these principles. Philosophers offer contractarian arguments in various ways and to various purposes, but they all share this fundamental conviction. As Rawls emphasises, at the heart of the social contract tradition is the notion of agreement.² Unless a political principle can be shown to represent the outcome of an agreement, its authority is unjustified and individuals are not obligated to follow it.

This methodology has a long pedigree and enjoyed a revival in 1970s. However, although it is widely used, the methodology itself receives very little attention. There are three main attitudes in philosophical discussions of contractarianism. First, philosophers might concentrate on specific contractarians.³ Secondly, philosophers might simply praise the methodology, and even argue that this is the only model of normative justification.⁴ Thirdly, philosophers might argue that this methodology is redundant or untenable.⁵ I believe that all three attitudes are inadequate. The first attitude fails to abstract from discussions on particular contractarians and study contractarianism from a more general perspective, while the second and third

¹ Anscombe (1981: 136).

² Rawls (2007: 159).

³ See, for example, Boucher and Kelly (1994) and Vallentyne (1991a).

⁴ Gauthier (1991a).

⁵ See Dworkin (1973), Pettit (2006) and Hooker (2002).

approaches only focus on one side and are not critical enough. A general and critical discussion of the contemporary development of this methodology is almost absent.⁶ The aim of this thesis is to fill this space and discuss the methodological value of contractarianism by comparing various contemporary contractarian models.

The question that interests me is whether contractarianism is a good approach to justifying the normativity of political principles. Here, normativity refers to normative reason. When something is normative, it gives us a reason for action in a particular way.⁷ Can contractarianism justify certain political principles as normative? What are the problems of the contemporary contractarian theories? For these problems, are they merely an outcome of applying this methodology badly, or are they unavoidable consequences of employing this methodology? In this thesis I will show that, by examining three different contractarian models, we can see that contractarianism, as a general methodology, has critical problems for justifying political principles that any theory based on this methodology cannot avoid.

A distinctive feature of contemporary contractarianism is that it presupposes a conception of practical reason in order to explain its source of normativity. Based on three different conceptions of practical reason, three contractarian models can possibly be developed, which are also the three contractarian models that appear in contemporary political philosophy: Hobbesian contractarianism, Kantian contractarianism and hybrid contractarianism.⁸ These three models have different

⁶ There are some monographs about the development of classical contractarianism, but not about contemporary contractarianism, such as Lessnoff (1986) and Riley (1982). Most of the general discussions are superficial and merely used for an introductory purpose, such as Darwall (2003b), Sayre-McCord (2000) and Hampton (1993a).

⁷ This definition of normativity is a fairly common view nowadays. See Finlay (2010).

⁸ I should clarify here my use of the term 'contractarianism', for there is no a standard way of using this term in current academic discussions. Stephen Darwall (2003b) uses 'contractualism' to refer to the strand of social contract theories which are influenced by Kant and emphasize agreement which could be reasonably accepted by everyone, and uses 'contractarianism' to refer to the strand of social

weaknesses, and I will further argue that if these weaknesses are understood against a wider background, we can see that these weaknesses are actually linked to the contractarian methodology. Hobbesian and Kantian contractarianism are vulnerable to the problem of priority because the former ignores reasonableness and the latter ignores rationality in their conceptions of practical reason. Although hybrid contractarianism avoids this problem by incorporating both rationality and reasonableness into the conception of practical reason, it suffers from another problem: it would be accepted only by a limited number of people. By studying these three models, we can see that contemporary contractarianism inevitably faces a dilemma: either it fails to prove that these political principles have the highest priority, or it fails to demonstrate that these political principles are generally acceptable. This dilemma can be attributed to the tension between rationality and reasonableness in our practical reason. As long as the contractarian methodology relies heavily on practical reason as the source of normativity, this dilemma between priority and generality is inevitable to contemporary contractarians. Once a philosopher employs this methodology, he will face this dilemma no matter how he employs it. Through a general study of contemporary contractarianism, a flaw of this methodology which is generally unrecognised by both contractarians and anti-contractarians alike can be made explicit.

contract theories which are influenced by Hobbes and emphasize agreement which could advance the rational interests of everyone. However, not everyone agrees with this label. For example, Geoffrey Sayre-McCord (2000) and Jean Hampton (1993a) use 'Kantian contractarianism' to refer to the former family and 'Hobbesian contractarianism' to refer to the latter family. Nicholas Southwood (2010) uses a distinction between Hobbesian and Kantian contractualism (Southwood 2010). Samuel Freeman prefers a distinction between an interest-based contract and a right-based contract (Freeman 1990).

I disagree with all the distinctions above since they overlook the third possibility of contractarianism, the hybrid model. In my usage, contractarianism is a broad term which can be used to refer to all the theories that use social contracts to justify political principles, regardless of whether this contract is rational or reasonable. Hence, unlike Darwall's usage, this term does not specifically refer to the Hobbesian model of social contract theorising. Under the label of contractarianism, I derive into three families: Hobbesian, Kantian and hybrid. I will discuss these three models in this thesis.

Before beginning, it would be instructive to define the object of my discussion and some key terms. The purpose of this introduction is to set the boundaries of my discussion. Since contractarianism is such a large school and this methodology is used by people in many different ways, it is impossible to discuss them all. I will first distinguish classical contractarianism and contemporary contractarianism. Then I will discuss which function of contractarianism I want to examine. I will also define political principles and explain why a justified political principle should satisfy two conditions. After that, I will briefly outline my whole argument and explain why the feminist critique of contractarianism, though it is well-known, is not relevant to my discussion.

1.1 A very brief history of contractarianism

Contractarianism has a long and distinguished history. Its roots can be traced as far back as Plato's Republic, where Glaucon presents it as a view of justice. It also appeared in political writings in medieval times and had its heyday in the seventeenth and eighteenth centuries.⁹ The popularity of contractarianism can be attributed to the Reformation, which strengthened the ideas of free individual choice and responsibility in moral thinking. At that time the majority of philosophical notables used the notion of the social contract to justify their political theories. These included Hugo Grotius, Thomas Hobbes, Samuel Pufendorf, John Locke, Jean-Jacques Rousseau and Immanuel Kant.¹⁰ However, contractarianism's popularity started to decline in the late nineteenth and early twentieth century, because of the rise of utilitarianism and Marxism. These substantial alternatives advanced their own

⁹ Gough (1957: 1).

¹⁰ See Grotius (2005), Hobbes (1994), Pufendorf (1991), Locke (1988), Rousseau (1987), Kant (1998), Kant (1991).

political theories along with their criticisms of contractarianism.¹¹ Nevertheless, contractarianism did not become a doctrine of purely antiquarian interest. Its popularity enjoyed a renaissance in the later twentieth century due to the publication of John Rawls' *A Theory of Justice*. Now, there are a number of political philosophers who develop their theories on the basis of this methodology, including David Gauthier, Gregory Kavka, James Buchanan, T. M. Scanlon, Brian Barry, Thomas Nagel, David Richards, and John Harsanyi.¹²

This is a very brief description of the development of contractarianism; from this we can see that contractarianism has a history of over two thousand years. It is unrealistic to expect that all contractarians throughout this period used precisely the same methodology. Indeed, with the exception of some very basic convictions, the structure of contractarian argument has changed significantly in its long history. For reasons of space, I cannot include all contractarian models in this discussion. I have chosen to focus on the contractarian theories that appeared in the latter part of the twentieth century. However, the characteristics of contemporary contractarianism can still be made explicit by comparing it with classical contractarianism. Classical contractarianism refers to those contractarian theories that appeared in the medieval period and flourished in the seventeenth and eighteenth centuries. By contemporary contractarianism, I mean those contractarian theories that appeared during the later twentieth century. Contractarianism underwent a massive change when it developed from classical contractarianism to contemporary contractarianism. This change can be seen in three dimensions.

¹¹ See, for example, Hume (1994) and Marx (2000).

¹² See Rawls (1999a), Rawls (1993), Gauthier (1986), Kavka (1986), Buchanan (1977), Scanlon (1998), Barry (1995), Nagel (1991), Richards (1971) and Harsanyi (1976).

First, the contractual relationship gradually changes from a relationship between the ruler and the ruled to a relationship among the ruled. At the beginning, the social contract was conceived as a bilateral contract between the people and the ruler. On the one hand, people elected a ruler and agreed to obey him. On the other hand, the authority of the ruler was also subject to certain limits. This understanding of social contract was prevalent in the medieval period, and was also adopted by Grotius and Pufendorf, and to some extent Locke. However, this understanding was succeeded by another understanding after the seventeenth century. To Hobbes, Rousseau, Kant, and the contemporary contractarians after them, the social contract was a multilateral contract among individuals living in the same society. These contractarians present a scenario in which all of these individuals discussed together and agreed upon certain kinds of political arrangement. The contractual relationship in contractarianism changed because the form of state changed from a monarchy to a constitutional state. In a monarchy, when the kings stood apart from and in obvious contrast to their subjects, it was natural and useful to think in terms of a contract which defined the relationship between them. But in a modern constitutional state, all people, including the rulers, are united under the terms of a constitution. When both government and citizens are only parts of society subject to the rule of a constitution, the focus of contractarians naturally changes to what this constitution should be. As Sir Ernest Barker argues, in contemporary contractarianism, ‘the one political contract—which unites us all (rulers and subjects alike) in terms of a constitution, and under the constitution—this one contract is enough, and it is the only contract’.¹³

Secondly, the hypothetical character of the social contract is strengthened. In the beginning, the social contract was taken to be an historical explanation of the origin of the state. Some medieval contractarians believed that, in the distant past, a group of people, living in a state of nature, really did agree to be bound by a contract.

¹³ Barker (1980: xiv).

Hence political principles bound us because of the explicit consent of our ancestors and our tacit consent. However, the historical character began to fade away along with the development of contractarianism. This is because there is obviously no evidence that such a historical event actually occurred. To quote Barker again, from a historical perspective, ‘society is not constituted, and was never constituted, on any basis of contract’.¹⁴ Hence more and more contractarians started to concede that their social contracts were actually hypothetical. Contractarians were no longer interested in the historical origins of the state, but wanted to provide a philosophical description of how individuals would make a social agreement under suitable circumstances.¹⁵ This model of social contract theorising appeared in the seventeenth century. But at that time contractarianism, such as that of Hobbes and Locke, still had a dual character. They presented a hypothetical story of how a group of individuals could agree upon a social contract, but also believed that there were historical events that supported their hypothetical contractarian model.¹⁶ After Rousseau and Kant, the historical character of contractarianism completely disappeared and, nowadays, all contractarians argue that there is no historical basis for thinking that any original contract was ever made. The social contract is merely a hypothetical test to evaluate whether or not the existing state is justified. This is why

¹⁴ Barker (1980: xiii).

¹⁵ Gough (1957: 244).

¹⁶ Hobbes’ theory is a reduction of human nature, and the state of nature is only a kind of fictional or hypothetical model (Hobbes 1994: 453-454) to show what society would become in the absence of state, and, inside that, the contract is a depiction of the regulations that might logically be accepted by people. However, in the Part III and IV of *Leviathan*, Hobbes also attempted to give a ‘Hobbesian’ interpretation to the history in the Bible for the sake of showing that the social contract that he presented really *existed* historically. For an interpretation of Hobbes as a hypothetical contract theorist, see Peters (1956: 158).

Similar dual features can also be found in Locke. Locke is normally understood as a theorist very different from Hobbes but who still argued that the social contract was a real historical event and that the obligations we have now are derived from this old contract. Locke himself tries to give such a history in sections 100-122 of the *Second Treatise*. However, on the other hand, Locke’s social contract theory has hypothetical character. For an interpretation of Locke as a hypothetical contract theorist, see Pitkin (1972).

Kant, for instance, asserted: 'If the law is such that a whole people could not possibly agree to it...it is unjust'.¹⁷

Thirdly, the account of political authority changes from voluntarism, which respects the choice of people, to rationalism, which supposes that their choices must follow certain definite rules. The idea of consent becomes gradually less important. Classical contractarianism usually takes voluntary consent to be the source of the normative power of social contract.¹⁸ A political principle is justified because it has the voluntary consent of the people who are subject to it. But what counts as voluntary consent? There are usually two accounts, explicit consent and tacit consent, yet both are problematic. It seems that even the most justified political principles cannot receive the explicit consent of every citizen. Moreover, even if such a consensus can be reached, it may be the product of coercion or misinformation. Tacit consent is not satisfactory either, since an unjustified state can also invoke tacit consent to justify its political authority. The account of tacit consent is too open-ended and depends on the interpretation of the state. A state can define tacit consent however it likes and people cannot defend themselves against a claim which has not been actually consented to by them.¹⁹ While both accounts are problematic, contemporary contractarians (and also some classical contractarians,

¹⁷ Kant (1991a: 79).

¹⁸ One of the examples is Hobbes. As we saw before, Hobbes' social contract is hypothetical. But because of its dual character, he also believed that the political obligation to obey a sovereign is based on the actual promise of individuals to an actual sovereign: 'every subject in a Commonwealth, hath covenanted to obey the Civil Law...And therefore obedience to the Civil Law is also part of the Law of Nature'. (Hobbes 1994: 175) Hobbes even argues that political obligation can only be derived from the actual promises of individuals. There is 'no obligation on any man, which ariseth not from some act of his own', hence our political obligation depends on 'what rights we pass away, when we make a commonwealth'. (Hobbes 1994: 141) Therefore, it seems that, to Hobbes, our obligations to political principles are not only justified by hypothetical consent. The actual consent is also a necessary condition.

¹⁹ Since my purpose is merely to contrast contemporary contractarianism with classical contractarianism, I will not examine the criticism of explicit and tacit consents in details here. For detailed criticisms of explicit and tacit consent, see Wolff (2006: 41-44), Simmons (1981) and Pateman (1985).

such as Kant) place less emphasis on voluntarism and look for other accounts of political authority. They argue that not all forms of consent can authorize political principles. Only rational (or reasonable) consent can authorize. This also means that, provided that a political principle is able to receive the rational consent of people, then it is justified even though it has not been voluntarily consented to by people. Contemporary contractarians place more importance on the idea of rationality (and reasonableness) than on the idea of voluntary consent. Unlike classical contractarians, who take consent itself to be the source of normative power, contemporary contractarians see consent as a heuristic device that reveals what political principles people are rational (or reasonable) to agree to.

By comparing classical contractarianism and contemporary contractarianism, the distinctive features of contemporary contractarianism become more evident. In the long history of contractarianism, different critiques of this methodology have appeared and contractarians have continued to revise this methodology. Nowadays this methodology is highly different from its original version. Classical contractarianism conceives of the social contract as an historical, voluntary contract formed between the ruler and the ruled. As we saw before, for different reasons, these three characteristics were gradually abandoned by contractarians. Currently, contemporary contractarians clearly define their social contract as a hypothetical, rational (or reasonable) contract formed among individuals.

1.2 Various functions of contractarianism

In the last section we saw the difference between classical contractarianism and contemporary contractarianism. However, in the contemporary age, the social contract has a variety of uses. It is crucial to clarify which function of contractarianism I am concerned with. First, contractarianism can have two functions: descriptive and evaluative. For descriptive contractarianism, the social

contract is used to explain why morality or political institutions would naturally emerge among human beings. According to Sayre-McCord, this approach is inspired by David Hume, who took for granted ‘neither a concern for morality nor any particular account of what people have reason to do or accept’.²⁰ It merely uses the social contract to make sense of the evolution of existing morality and political institutions. Descriptive contractarians argue that, through social contract argument, the rationale behind morality and political institution can be made explicit.

This approach is normally adopted by game theorists who use mathematical models to explain how human beings interact with one another and finally form social conventions, which are the implicit social agreements on morality and political institutions. Representatives of this approach are Ken Binmore, Brian Skyrms, and to some extent Jean Hampton.²¹ To these theorists, the social contract refers to the social conventions that act as the basis of morality and political institutions. Insofar as people continue to support these social conventions, these conventions continue to exist and comprise morality and political institutions. However, these social conventions are not formed arbitrarily. They are created by the interactions between people and these interactions can be articulated clearly by game theory. These theorists believe that, through game theory, we can see how people coordinate their behaviour in equilibrium in the game of life. This equilibrium then forms a set of common understandings among people, and this is where the social conventions come from. Hence, the descriptive contractarians are concerned with what morality and political constitutions are. The nature of these social practices is the social conventions that evolve from the rational interactions between individuals. I would like to note that this approach to contractarianism is merely descriptive. Even if a state is based on this social convention, it does not mean that this state is a justified

²⁰ Sayre-McCord (2000: 263). See also Binmore (2005: 3-4).

²¹ Binmore (2005), Skyrms (1996) and Hampton (1986).

or legitimate regime. It merely shows that contractarianism successfully explains how a state is generated.

This approach to contractarianism is not the kind of contractarianism that I will discuss in this thesis. The approach that interests me is the approach that uses contractarianism in an evaluative way. In evaluative contractarianism, the social contract is used to justify certain moral and political principles that can impose obligations on individuals. It is concerned with what moral principles or political principles ought to be. Contractarians first imagine a hypothetical situation. A group of people discuss with one another under appropriate circumstances and agree upon some principles. This hypothetical contract can then be used as a testing device of the justifiability of the existing moral customs or states. The moral or political principles which are defined by this hypothetical contract have such normative authority that real people are obligated to follow them, despite the fact that they did not actually agree to them. Therefore, the social contract can be used to perform two distinctive functions. It can be used to describe the nature of morality or political institutions, and can be used to evaluate the justifiability of a given moral outlook or political institution.

However, we should be aware that contractarians usually do not have only one ambition. They can attempt to explain some existing moral or political principles, but may also want to justify these moral or political principles. For example, James Buchanan uses his contract theory to make sense of some social relationships (such as slavery), but he also argues that a justified law should represent a mutually advantageous agreement.²² Scanlon says that his project is to give a ‘phenomenologically accurate’ account of moral motivation, but he also believes that

²² Buchanan (1977: 56).

people ought to follow principles that could not be reasonably rejected.²³ Hampton even explicitly claims that her contractarianism has two functions: explanatory and justificatory. ‘The contractarian, while explaining how any existing state is a human creation and how it is a product of human convention, must leave behind history and posit a process of state creation that will reveal what reasons make that creation rational’.²⁴ Because of some contractarians’ two-fold ambitions, readers might confuse these two functions of contractarianism. Nevertheless, I believe that these two functions are independent and I will focus only on the evaluative function of contractarianism in this thesis.

Furthermore, even if it is used for justificatory purposes, the social contract can be used to justify either moral principles or political principles. Both of these two kinds of principles are normative in nature, but the scopes that they govern are different. In general, the scope of moral principles is much larger than that of political principles. Moral principles represent general regulation of conduct. They purport to set forth conditions under which an action is right or wrong or something is good or bad. In Socrates’s words, it is about ‘how we ought to live’. These principles govern various aspects of human life, such as how we should treat one another, how we should treat the environment and animals and how we should treat ourselves. Compared with moral principles, the scope of political principles is more limited. Political principles are also a kind of moral principle since they also regulate conduct. However, these principles are applied only in basic political institutions, such as the political constitution, the legal system, the system of economic redistribution and the market.²⁵ They govern only the political aspect of human life, that is, how the state should relate to individuals. For example, political principles define how a ruler can

²³ Scanlon (1998: 155).

²⁴ Hampton (1986: 185).

²⁵ This definition of a basic social institution is influenced by Rawls’ definition of basic structure. See Rawls (1999a: 6-10).

justifiably rule over citizens. They also define what laws citizens are obligated to follow and how citizens ought to behave in the public domain. In general, both moral and political principles share the same nature, but the latter has more limited concerns. In this thesis, contractarianism that justifies political principles will be my main focus.

Some contractarians, such as Scanlon, Gauthier and Richards, use the social contract to justify morality (moral contractarianism). They are more ambitious and aim to use the social contract to provide a foundation for morality. But some contractarians, such as Rawls and Barry, have more limited ambitions and clearly concede that they are only interested in justifying a normative theory of political institution (political contractarianism). At first glance, it seems that the former group of contractarians are irrelevant to my discussion. However, since political principles are also a kind of moral principle, the former group of contractarians cannot avoid the issue of political morality. As some moral contractarians acknowledge, the moral principles that they justify do have political implications.²⁶ Hence, although my concern is political contractarianism, the political implications for some moral contractarian theories will not be overlooked.

Hence, by illustrating the distinction between descriptive and evaluative and the distinction between moral and political, the kind of contractarianism that I will discuss in this thesis can be clarified. The contemporary contractarian theories that I will discuss are political contractarian theories which aim at justifying political principles.

²⁶ Scanlon (1998: 6).

1.3 The nature of political principle

It will be helpful to clarify what political principles mean and why the justification of principles is fundamentally important in political philosophy. People inevitably live together and form a state. The state is influential because it can coerce people and shaper people's lives. For example, people have to pay taxes, their careers and wealth are affected by the economic policies, and what they are free to do is also defined by the law of that state. The coercion of a state even involves matters of life and death, since a state conducts wars and uses deadly methods to punish dangerous lawbreakers. Given that a state has this much control, why should we accept such an authority? As Jean Hampton argues, political authority should be able to give reasons to its subjects so that those subjects understand that the coercion is rightful and legitimate, and not merely take coercion for granted.²⁷

Political principles are the conditions that justify this political authority. They represent a set of normative standards detailing what a state should be and how coercive political power can be justifiably exercised. Given that a state rules in accordance with these political principles, its political authority is justified and it can justifiably enforce rules upon its citizens as well. One should not confuse political principles with existing laws here. Indeed, political principles are only abstract normative principles and do not refer to any existing laws. Existing laws may fail to conform to these principles, but this merely means that existing laws are not satisfactory and ought to be revised. Moreover, these political principles also define what political obligations people have to their state. If the political authority is justified, then people will have a political obligation to obey it.²⁸ The laws or

²⁷ Hampton (1997: 4-5).

²⁸ This relationship between political principles and obligations may be doubted by some theorists. For example, Cynthia Stark argues that even if political principles are justified, it only means that

commands from this political authority should be taken as a reason for action. Political obligations stand for what people should do under a justified political authority. As John Horton argues, a justified political obligation represents a proper relationship between the state and its citizens.²⁹ Hence, political principles generally define two things: how a state can justifiably rule its citizens and what obligations citizens have to the state.

But how can we evaluate a political principle? There are two properties that a political principle must possess. First, a political principle must be general, which means that it must be justified to each citizen in society publicly. Political principles can be justified on cultural grounds or religious grounds. However, the authority of these political principles would only have limited acceptability because only specific cultural groups or religious groups would acknowledge their authority. This goes against our common thought about political authority, for we usually think that all citizens in society generally have an obligation to follow the commands of political authority. The authority of political principles is unconditional and objective. People should have reasons to acknowledge the authority of political principles, irrespective of who they are and what group they belong. This feature of political principles is also accepted by contractarians since they also believe that, even though citizens in a society are highly diverse, citizens should still generally accept the same social contract and this social contract can form the basis of social cooperation.³⁰

Secondly, a political principle must be overriding, which means that the reason to follow political principle should be accorded lexical priority (or very strongly

people have a moral obligation to follow it but political obligations depend on voluntary consent. See Stark (2000).

²⁹ Horton (1992: 13).

³⁰ Contractarians generally hold this viewpoint. See Gauthier (1986: 19), Rawls (1999a: 114) and Rawls (2001:32-33).

weighted priority) over other reasons. Citizens may have various reasons for doing something that political principles forbid. For example, they may have different projects and relationships that they want to pursue. These projects and relationships may be crucial to citizens and can be reasons for citizens to challenge the political authority. A justified political principle must prove that the reason to follow political principles can override or outweigh other competing reasons. The laws and commands from this authority should be authoritative in determining what citizens ought to do. This is why Joseph Raz argues that the demands of a political authority should operate as exclusionary or pre-emptive reasons.³¹ If citizens agree with a political principle, but still place it below other commitments in their lives, then this principle exists in name only, for it does not possess the supreme authority that we think a political principle should possess. This is also the ambition acknowledged by many contractarians for they normally aim at proving that the political principles justified by their social contract can override other commitments in people's lives. As Rawls emphasises, '[citizens] desires and aspirations are restricted from the outset by the principles of justice which specify the boundaries that men's system of ends must respect'.³²

Hence, a justified political principle must meet these two conditions:

- (i) Generality: It must be generally acceptable to the citizens who are subject to its ruling*
- (ii) Priority: It must explain why it can override other commitments in a citizen's life*

³¹ Raz (1979a: 23-33).

³² See Rawls (1999a: 27). Similar point can also be found in Gauthier (1986: 2) and Scanlon (1998: 160-168).

In general, the first condition is a requirement of *scope*, to the effect that the political principle can be generally accepted by people as a justified, normatively binding principle. The second condition is a requirement of *status*, to the effect that, to those people who accept the political principle, this principle has deliberative priority over other considerations in practical reasoning.

These two conditions are necessary since they are closely related to our intuitions of what a justified political authority should be. As H. L. A. Hart says, a political authority should be binding on all subjects and should be peremptory.³³ While the first nature refers to the general acceptability of political authority, the second refers to the priority of political authority. If a political principle is general but not overriding, then although citizens generally accept the political obligations, they may still challenge the authority of this principle because of other commitments they have in their life. If a political principle is overriding but not general, then although citizens who accept the political obligations will respect the supreme authority of this principle, these obligations will not be accepted by the majority of citizens in the society. Therefore, a political principle that fails to satisfy either of these conditions is merely a coercive power that commands citizens in an unjustifiable way. This section offers only a brief description of these two conditions. In Chapter 2, I will discuss these two conditions in more details and study what kind of contractarian models can satisfy these two conditions.

1.4 Method and outline

In section 1.1 and 1.2 I identified the kind of contractarianism that I am concerned with: contractarian theories which aim at justifying political principles. In section 1.3

³³ Hart (1982: 23-68).

I suggested two criteria to evaluate whether a political principle is justified. The aim of my thesis is to use the latter to evaluate the former.³⁴ In the next six chapters, I will discuss whether or not contractarian theories can provide a satisfactory justification for political principles.

In Chapter 2, I will discuss contractarianism as a general methodology before I begin to discuss different contractarian models separately. Although contractarianism had a revival in the second half of the twentieth century, it suffers from a number of famous critiques. I will first discuss the virtues of this methodology so that we can understand why so many philosophers are preoccupied with it. Then I will discuss the ‘standard critique’ of this methodology, which is to say, that a hypothetical contract does not have any normative force. I argue that the normativity comes from the conception of practical reason behind contractarianism. Provided that practical reasoning is governed by certain conceptions, and these conceptions justify certain political principles, people will have obligations to these political principles even though they have not consented to these principles.

From the discussion in Chapter 2, we should gain a better understanding on the structure of the argument of contemporary contractarianism. Contractarian argument consists of three elements: the conception of practical reason, the characteristics and circumstances of hypothetical contractors, and the hypothetical agreement. The first

³⁴ Since I define political principle as the object of my thesis, my thesis will not consider G. A. Cohen’s critique of constructivism (although Cohen uses the label of constructivism, most of the constructivists under this label are contractarians). Cohen’s critique is that constructivists are asking a question of what the optimal rules of social regulation are, how our common social life should be governed. But this is different from the question of justice, for justice should be insensitive to existing social facts. Hence, we should distinguish between principles of justice and rules of regulation. My thesis will not consider Cohen’s critique because my thesis is *exactly* about rules of regulation in Cohen’s definition, and most of the contractarians that I will discuss are concerned with rules of regulation as well. Hence, the questions that I am going to discuss are irrelevant to the question of whether or not constructivist principles reveal a ‘pure’ conception of justice. For Cohen’s critique, see Cohen (2008: 274 -343).

element determines the content of the other two. In Chapter 3, I will argue that, based on different conceptions of practical reason, different contractarian models can be developed. There are two aspects of practical reason: rationality and reasonableness. Based on the conception of rationality, contractarians such as Gauthier justify a social contract that could advance the rational interests of everyone. I call this model Hobbesian contractarianism. Based on the conception of reasonableness, contractarians, such as Scanlon justify a social contract that could not be reasonably rejected by anyone. I then further argue that, apart from these two models which are commonly recognised, there is also a third possible contractarian model, hybrid contractarianism. Hybrid contractarians, such as Rawls, assume that people are both rational and reasonable, and thus a social contract should be both rationally and reasonably justified. These three models represent three approaches to employing the contractarian methodology to justify political principles.

The next four chapters are then case studies of these three models. In Chapter 4, I will discuss Hobbesian contractarianism. Hobbesian contractarians construct their theory on the basis of a formal conception of rationality and assume that hypothetical contractors are rational utility-maximizers. Then political principles should represent a hypothetical contract which could be mutually advantageous to every rational contractor. I argue that this contractarian model can satisfy the condition of generality but fails to satisfy the condition of priority. It can satisfy the condition of generality because the formal conception of rationality is so weak that it can be generally accepted by everyone, so the political principles which are rationally justified should also be generally justified. However, it fails to satisfy the condition of priority because it overlooks the conception of reasonableness, which is another aspect of practical reason. Reasonableness may justify obligations which are incompatible with these political principles. Since reasonableness is equally important as rationality in practical reason, actual people have no reason to prefer complying with these political principles to violating them.

I will argue that Kantian contractarianism makes a similar mistake in Chapter 5. Unlike Hobbesian contractarianism, Kantian contractarianism relies on a conception of reasonableness. Hypothetical contractors are reasonable people who mutually respect each another by behaving in a way which could be justifiable to each other. Political principles should represent a hypothetical contract which could not be reasonably rejected by anyone. Again, this contractarian model can satisfy the condition of generality, but it fails to satisfy the condition of priority. The conception of reasonableness is a generally accepted conception of practical reason, so Kantian contractarianism, which is based on this conception of practical reason, should be able to satisfy the condition of generality. However, it fails to satisfy the condition of priority because, in opposition to Hobbesian contractarianism, it overlooks the conception of rationality, which is another aspect of practical reason. If political principles are irrational, then even though people generally accept these principles, they may refuse to acknowledge the supreme status of these principles.

In the failure of these two models, we can see that the key to satisfying the condition of priority may be to include both rationality and reasonableness into the conception of practical reason. This is exactly what Rawls seeks to do in his contractarian theory. Chapter 6 will be devoted to discussing the earlier version of Rawls' contractarian theory. In his theory, hypothetical contractors, free and equal persons, are assumed to possess two moral capacities, which correspond to rationality and reasonableness respectively. These contractors are willing to behave in a justifiable way, but are also concerned with several 'highest-order interests'. Thus the social contract they will make should be both rationally and reasonably justified; it represents regulation which could be publicly justified, but also represents an effective way to advance the highest-order interests of individuals. In this chapter I point out the possibility that these highest-order interests may conflict with each other, but I also suggest that Rawls' contractarian model can still satisfy the condition of priority by adding certain presuppositions. Nonetheless, although it can satisfy the condition of priority, it fails to satisfy the condition of generality. For, in

order to show that rationality and reasonableness lead to the same political principles, Rawls adopts a substantial conception of rationality which is not generally acceptable. Due to this flaw, the condition of generality becomes the Achilles' heel of hybrid contractarianism.

Rawls recognises this flaw, and makes a political turn in his later writings. Chapter 7 discusses whether this can rescue Rawls' theory. The later Rawls makes two changes, one to admit that the application of his theory is only limited to liberal democratic societies, another to emphasise that the assumptions of his theory are so freestanding that they are compatible with most of the comprehensive doctrines in the liberal democratic societies. I argue that, even if we lower the standard of generality and concentrate only on liberal democratic societies, Rawls' social contract is still not generally acceptable. In spite of being freestanding, the assumptions of Rawls' contractarian theory are actually more substantial than Rawls acknowledges. These assumptions presuppose that people are 'strongly reasonable citizens' who are willing to take political conceptions as the only ground of justification for political principles. However, even in a well-ordered society, which is the most idealist form of liberal democratic society, people may still not be 'strongly reasonable citizens' due to the freedoms allowed in their society. I will argue that citizens in the well-ordered society are generally 'weakly reasonable citizens', so the assumptions of hybrid contractarianism will only be accepted by a small group of 'strongly reasonable citizens' in the well-ordered society.

After we examine all these three models of contractarianism, we can discover a general weakness of contractarianism as a methodology. These three models are three possible approaches to employing this methodology. When all of them have problems, we must ask whether there is something wrong with this methodology. The distinguishing feature of contemporary contractarianism is that it relates political principles and practical reason by hypothetical contract. But this also means that

contractarianism inevitably has to rely on practical reason, or else it is unable to explain the source of normativity. However, if contractarians rely on a conception of practical reason, then they have to choose between generality and priority. Either they assume rationality or reasonableness as the conception of practical reason, or they include both rationality and reasonableness into the conception of practical reason. In the former case, contractarianism loses the virtue of priority. In the latter case, they lose the virtue of generality. Contractarianism is caught on the horns of a dilemma: once political philosophers adopt this methodology, they cannot avoid it.

1.5 Contractarianism: a patriarchal theory?

My aim in this thesis, as I have said, is to provide a general discussion of contractarianism as a methodology, but I have to admit that my discussion cannot cover all contemporary critiques. While contractarianism is one of the most popular approaches of justification in contemporary political philosophy, it has also been the subject of various critiques. For example, Carole Pateman famously criticizes contractarianism as an unsatisfactory political theory because patriarchal relationships are implicit in it.³⁵ This is a well-known feminist critique.³⁶ However, for reasons of space, it is impossible for me to deal with this critique in detail. Moreover, since I am concerned with the normativity of contractarianism, which is about its binding power, whether contractarianism excludes or disadvantages a particular group is not central to my concern. Hence, I will now discuss only briefly this feminist critique.

³⁵ Pateman (1988). See also Pateman and Mills (2007).

³⁶ See Coole (1994: 204-207), Lessnoff (1986: 107-108), Hampton (1993a: 492) and Sayre-McCord (2000: 267).

My point is that classical contractarianism may be vulnerable to Pateman's critique, but her critique surely does not pose a problem for contemporary contractarianism. Pateman's argument is that a sexual domination underlies all social contract models and this sexual domination is always ignored by political philosophers. 'The sexual contract is a repressed dimension of contract theory, an integral part of the rational choice of the familiar, original agreement'.³⁷ Contractarian theories cannot avoid this sexual domination because they presuppose that only individuals who are endowed with certain capacities or attributes can enter into the social contract. However, these capacities or attributes are usually defined through a very masculine approach. At the end, only men are qualified to participate in the social contract. 'Women are *excluded* from the original pact. Men make the original contract'.³⁸ The role of women in civil society is only as the property of men, subjected to the rule of men. Women are part of the civil society, but they do not enjoy the same protection and freedom that men do. Therefore, Pateman argues, even though contractarianism has an appearance of equality and justice, it is actually unsatisfactory because it allows a patriarchal relationship between men and women to exist in society. Beneath the equal social contract is the unequal sexual contract.

Men's dominance over women, and the right of men to enjoy equal sexual access to women, is at issue in the making of the original pact. The social contract is a story of freedom; the sexual contract is a story of subjection. The original contract constitutes both freedom and domination. Men's freedom and women's subjection are created through the original contract.³⁹

Classical contractarians, as Pateman observes, really overlook the status of women in their social contract. For example, Locke seems to think that women, unlike sons,

³⁷ Pateman (1988: iv).

³⁸ Pateman (1988: 5).

³⁹ Pateman (1988: 2).

can never emerge from the ‘protection’ of men.⁴⁰ However, this critique does not apply to contemporary contractarianism. As we will see later in Chapter 2, in contemporary contractarianism, the qualification of contractors is that they are capable of engaging in practical reasoning governed by certain conceptions, such as rationality or reasonableness. The capacity for practical reason is so common that every human being should possess it. The criteria of qualification do not have the kind of masculine inclination described by Pateman. For example, Scanlon assumes that every contractor has a capacity for reasonableness. In his contractarian model, contractors are reasoning creatures that have a capacity to recognise, assess, and be moved by reason.⁴¹ They can give reasons to justify their acts and respect each other as creatures that are capable of asking for justification. This capacity is so common that, as Scanlon believes, all human beings possess it. If having this capacity is the qualification of being included in the social contract, then all women should be capable of entering into the contract. Indeed, many contemporary contractarians are also aware of this underlying patriarchal relationship. Even Rawls, who is criticized for only allowing heads of families to enter into social contract and ignoring other family members,⁴² also clarifies that principles of justice protect the basic liberties of all family members.⁴³ Provided that a person has two moral capacities, that is, they are rational and reasonable, they are in the scope of the social contract and protected by justice as fairness, regardless of gender. Rawls believes that this scope includes ‘the overwhelming majority of mankind’.⁴⁴ While the criteria of qualification in

⁴⁰ Pateman (1988: 94).

⁴¹ Scanlon (1998: 23).

⁴² Pateman (1988: 41-43) and Okin (1989: 93-97).

⁴³ Rawls (1997: 597).

⁴⁴ Rawls (1999a: 443).

contemporary contractarianism are so low, women will always be included and fairly treated in contemporary contractarian models.⁴⁵

Pateman might argue that, even though the criteria are weak, many women may still fail to qualify because they are socially and educationally disadvantaged. This reply would be over-cautious because the criteria I have mentioned are so weak that people who were not properly educated or reared should still be able to fulfil them. Even if women were so unfortunate that they did not have the chance to realise their capacities for practical reason, they could not be excluded from the social contract. For contemporary contractarians argue that people who have the *potential* to be capable of assessing these rules of practical reason must be included in the social contract. An individual may not have the chance to realise their potential, yet they can still enter into social contract given that they have the potential. As Rawls says, 'only scattered individuals are without this capacity, or its realization to the minimum degree, and the failure to realise it is the consequence of unjust and impoverished social circumstances, or fortuitous contingencies'.⁴⁶

Hence, in light of the weak criteria of qualification, I believe that Pateman's feminist critique cannot apply to contemporary contractarianism. In contemporary contractarian theories, women are never the property of men because they are participants in the social contract. What should worry contemporary contractarians more is the relationship between the hypothetical contract and these people. They can

⁴⁵ Kantian and hybrid contractarianism do not need to worry about this critique. But Pateman might still argue that Hobbesian contractarianism, represented by Gauthier, is vulnerable to her critique because, in his assumptions, contractors not only have to be rational, but also have to possess bargaining power. If women do not have enough bargaining power that can bring advantages to other participants, then they may be excluded. However, Gauthier seems to think that the bargaining power of women is not much weaker than that of men. The illusion that women would be disadvantaged is only because of some ideological and historical factors. Actually women should not have any problem in entering into social contracts. See Gauthier (1986: 230-231).

⁴⁶ Rawls (1999a: 443).

Chapter 1 Introduction

be present with a very beautiful hypothetical contract story, but how can this story bind people or define what people ought to do? This question will be addressed in the next chapter.

Chapter 2 Why Contractarianism?

2.1 Introduction

Twenty years ago, David Gauthier wrote an article ‘Why Contractarianism?’ explaining why contractarianism is an attractive methodology in the contemporary age. This article is valuable since there have been few methodological discussions of contractarianism before or since. Nevertheless, it is unsatisfactory because Gauthier’s account of contractarianism’s appeal is not comprehensive enough, and he does not respond to some commonly known critiques of contractarianism. For example, some philosophers, such as Ronald Dworkin, doubt the normative force of the hypothetical contract, but Gauthier does not take this critique seriously.¹ Before I begin my discussion of various contractarian models, I should deal with this unanswered critique and explain, from a methodological perspective, why contractarianism is still worth taking seriously today. Therefore, this chapter picks up Gauthier’s task again after twenty years.² The ambition of this chapter is to demonstrate that contractarianism is a useful method for justifying normative principles. I will provide a more comprehensive discussion of the strengths and weaknesses of this methodology.

¹ See, for example, Brudney (1991), Dworkin (1973) and Hampton (1993b).

² However, it does not mean that I am going to take Gauthier’s philosophical position, such as the ‘moral-error’ meta-ethical position. In ‘Why Contractarianism’(Gauthier 1991), Gauthier holds a ‘moral-error’ position that denies any objective moral constraints and moral fact. He argues that most of our moral intuitions are unreliable and hence moral principles are unjustified unless they have rational foundation. However, not all contractarians take this philosophical position. Some contractarians, like Scanlon and Nagel, disagree with this ‘moral-error’ standpoint. Other contractarians, like Rawls, even want to avoid this kind of meta-ethical controversies. Hence, I believe that contractarianism can be a freestanding methodology and does not necessarily connect with any meta-ethical position. Contractarian can have different understandings of what the nature of morality is, but still use the same methodology (hypothetical contract) to justify normative principles.

I would like first to define contractarianism. Here contractarianism refers to the normative political theories which use hypothetical contracts as a method for determining the nature and content of political principles. 'This doctrine of justification, tying reasons in interaction to the test of hypothetical agreement, is the core of...contractarianism'.³ There are two central features of hypothetical contract: an idealised choice situation and a hypothetical agreement among people. Contractarians first define an idealised choice situation in which there are no sovereigns, laws, courts, established property rights or contract.⁴ They then ask us to imagine that we have been placed in this ideal choice situation. In order to escape from the chaos, we would have to create political principles to govern our society. After deliberation, we would come to an agreement on certain political principles. Contractarians argue that these political principles (and our obligations to these principles) are justified, and use these political principles to evaluate existing social institutions. As Gauthier says in his article, the main task of the hypothetical contract is to establish an evaluative criterion: 'Hypothetical agreement provides a test of the justifiability of our existent practices'.⁵ This is the general form of contractarian argument.

Although contractarians disagree about the ideal choice situation that should frame the choice of political principles, they share this general framework of contractarianism. Some assume that people in an idealised choice situation would have strong moral motivation, whereas some assume that people would be wholly self-interested. Different assumptions of the idealised choice situation lead to

³ Gauthier (1997: 135).

⁴ Here 'idealized' does not mean those people who are in the idealized choice situation are perfect and virtuous. The term only means that, at this stage, some of their morally relevant characteristics are sharpened prominently, and, on the contrary, some of their irrelevant characteristics are excluded. So 'ideal beings' are human beings who have certain prominent characteristics. The purpose of idealization is to eliminate the influence of factors we consider irrelevant to the decision making of hypothetical agents.

⁵ Gauthier (1991a: 101).

different hypothetical agreements. From the variety of contractarian models, one should notice that contractarianism is a methodology which is compatible with a diverse range of conclusions.⁶ The variety is so large that it is difficult to see them as belonging to the same philosophical camp. Someone may then think that contractarianism is only an empty method which can be used by any political theorists.⁷ Anyone who has designed an idealised choice situation and proposed a hypothetical agreement can be called a contractarian. However, such a broad definition fails to explain why so many contractarians really take this philosophical position seriously. For example, even though his hypothetical contract story was subject to many critiques, Rawls never gave up the label of contractarian throughout his life and believed that it was a distinctive philosophical position.⁸ It seems that contractarianism has some special philosophical commitments apart from this general framework.

I believe that contractarianism has certain special characteristics and these characteristics differentiate contractarians from other political theorists. These characteristics can also be interpreted as the virtues of this methodology which explain why so many theorists are attracted to this approach. The distinctive virtues of contractarianism will be discussed in the following section. After this section, I will then discuss a common issue raised by critics of contractarianism, which is the relationship between the hypothetical contract and the real world.

⁶ Here contractarianism does not commit one to any particular meta-ethical theory about the nature or ontological status of injustice. It is compatible with the view that injustice is an irreducible, non-natural property, but it is also consistent with a meta-ethics that views attributions of injustice as ways of projecting one's disapproval onto actions. However, it does not mean that contractarianism does not have any meta-ethical implications. For example, contractarian principles must be possible to be 'chosen' by human beings. Hence the principles must be *stance-dependent*, that is, they must depend on an intentional psychological state (a stance), such as a belief or a conative or an affective attitude. For this meta-ethical implication of contractarianism, see Milo (1995: 191-192).

⁷ Kymlicka (1993: 186-187).

⁸ Freeman (2007b: 36).

2.2 Three virtues of contractarianism

Contractarianism has had its heyday, but there are still a number of contemporary theorists who continue to use this methodology. This phenomenon should not be ignored. Nor should we think that these theorists have coincidentally chosen the same method. There may well be some particular virtues inherent in this methodology, which are the key to explaining why it has been chosen. In this section, I will try to show what these virtues are and explain why contractarianism is so prevalent in political philosophy.

2.2.1 The virtue of naturalism

I can think of three virtues of contractarianism, which also seem to be the reasons why contemporary contractarians endorse this approach. First, contractarianism does not justify political principles by strange, non-natural properties or objects; nor does it credit human beings with what J. L. Mackie calls ‘magical power’ which is capable of discerning some moral truths out there’.⁹ Instead, it takes the normative principle as a ‘contract’, a product of human volition, that we commend to the extent that we would choose it when we are placed in an idealised choice situation. This characteristic is a virtue because it is compatible with the scientific, secular outlook in the modern world.

Here it is important to note the difference between the modern outlook and the ancient outlook. In the ancient times, people usually conceived of the world as purposely ordered.¹⁰ Everything was subjected to an order that different elements in

⁹ Mackie (1990: 38-42).

¹⁰ Taylor (1975: 6).

creation expressed and embodied a certain order of purposes. This order gave all beings innate purposes which would guide what they ought to do. The way people treated each another was somehow inscribed in the universal realm. Since this purposive order was independent of people's subjective will, people could only recognise their given purposes and pursue them faithfully.

This outlook changed radically following the Enlightenment. Because of the emergence of modern science, represented by the scientific theories of Newton and Galileo, people began to understand the order of the world as a mechanistic, non-purposive order. This is, in Max Weber's words, the 'disenchantment of the world', which is the feature of modernity. 'The new notion of objectivity rejected the recourse to final causes; it was mechanistic in the sense of relying on efficient causation only'.¹¹ In the new world order, people are not given any objective, final purpose of life.¹² The natural scientific order no longer tells people what they ought to do. The purposes of life do not exist objectively, but rather are determined by the subjective will of individuals. Therefore, unlike the ancient worldview that the whole world was governed by an independent purposive order, in the modern age, the factual world is explained by science, but science says nothing about what people *should* do.

This change in outlook influences the justification of political principles. The ancient approach to political justification took the political order to be a part of the cosmic order; thus a justified political principle was independent of what people thought. It

¹¹ Taylor (1975: 10).

¹² In a more accurate wording, it does not mean that no one conceives the world order as purposive during or after Enlightenment. Indeed, a purposive interpretation of history is still prevalent in the writings of Enlightenment thinkers, such as Kant (See Kant 1991b: 108-109). However, the purposive order in their mind is no longer an independent world order which exists regardless of the will of individuals. Rather, this purposive order comes from the projection of the wills of individuals. That is to say, the purposive order is not an order which is discovered by individuals, but rather a way of individuals to conceive the world.

could be justified by an external metaphysical order which was indifferent to the subjective will of the individual. However, this approach encounters severe difficulties in the modern world. The assumption that there is some external metaphysical truth out there which can be discovered by us and can guide our actions is incompatible with the scientific, natural worldview.

Hence contractarianism became the mainstream of political theory after the seventeenth century, since the justificational approach to contractarianism does not rely on any metaphysical properties which are indifferent to the subjective will of individuals. Contractarianism justifies principle by showing that it is the constraint which would be chosen by us in an idealised choice situation. In this approach, political principles come from the creation of individuals which represent the disposition of their subjective wills.¹³ Political authority is not something that can be derived from some sort of innate authority possessed by a set of supposedly superior persons, nor something that is derived from God. Instead, political authority should be a creation of the people who constitute it. Thus Michael Oakeshott is right to call contractarianism a doctrine of 'will and artifice'.¹⁴ Contractarians take the subjective will of people as the starting point for reflection on society. They deny 'an independent realm of moral facts and a special faculty of reasoning to ascertain them'.¹⁵ It is this naturalistic virtue which attracts some theorists to adopt this

¹³ Indeed, the feature of 'anti-natural authority' of the social contract can even be traced back to the earlier history of political thought. One of the examples of this is the 'nature and convention' debate among the Greek philosophers. As Sir Ernest Barker observes, early Greek philosophers normally had two approaches to explaining the existence of law, one accepted their own laws as unalterable law by nature, if not divine; the other argued that law was just a product of customs and conventions, the existence of political authority being based on the consent of ruled. The sophists of the latter camp, though they did not adopt contractual language, had very similar ideas on the formation of political organization to the modern social contract theorists. For a discussion of distinction between the 'nature' and 'convention' schools, see Barker (1960: 53). For the interpretation that the idea of consent-based political authority is actually implicit in the 'convention' school, see Kahn (1981).

¹⁴ Oakeshott (1975: 25).

¹⁵ Freeman (1991: 285). Although Freeman is only talking about the feature of Scanlon's contract theory here, it is also the common feature of contemporary contractarianism.

Chapter 2 Why Contractarianism?

approach, for example, David Gauthier, who believes that a normative constraint cannot be independent of the subjective will of people.

If, independent of a person's actual desires and aims, there were objective values, and if, independent of their actual purposes, they were part of an objectively purposive order, then we might have reason to insist on the inadequacy of the [ancient approach.]...But the supplanting of teleology in our physical and biological explanations closes this possibility, as it closes the possibility of religious explanation.¹⁶

Hence, contractarianism is an attractive approach to justifying political principles because it justifies principles in a way which is compatible with a modern scientific worldview, without relying on any questionable external metaphysical assumptions. No wonder some contractarians even believe that their approach can provide the 'only plausible foundation' of normative principle in the modern world.¹⁷

¹⁶ Gauthier (1991a: 98). Similar emphasis can also be seen in Rawls (1999a: 398), Rawls (1980: 350), Hampton (1986: 273) and Barry (1995a: 5).

¹⁷ Gauthier (1991a: 91).

2.2.2 The virtue of public justifiability

Secondly, contractarianism shows a respect for the choices of rational/reasonable individuals. Contractarians emphasise that political authority is unjustified unless it represents an agreement which is publicly acceptable to everyone in a society. This characteristic shows a basic normative commitment of contractarianism, that is, all persons should have equal moral standing. No-one is naturally superior or inferior to another. If we all have equal moral standing, then no-one, or no group, could have authority over others unless this authority is acceptable to others. An unjustified authority imposed on an individual represents disrespect for the freedom of this individual. As Gough points out, ‘the ultimate *raison d'être* for the contract theory, all through its history, has been to reconcile the apparently conflicting claims of liberty and law’.¹⁸

It does not mean that contemporary contractarians believe that a justified principle must be based on the voluntary consent of people, since sometimes the voluntary consent of people may not represent a rational/reasonable choice. Indeed, as I mentioned in Chapter 1, contemporary contractarians rarely take voluntary consent as the source of normativity.¹⁹ Nevertheless, contemporary contractarians still insist on the importance of public justifiability. Since a political principle frequently has a significant impact on the lives of people, contractarians argue that such an influential rule cannot be imposed on people arbitrarily. Even though it is unrealistic to show that a principle is a wholly voluntary scheme, contractarians believe that, at least, a principle should show that rational individuals would voluntarily accept its ruling. As John Rawls says, a political society should not be taken as a straightforwardly voluntary scheme. Nevertheless, a political society which is governed by

¹⁸ Gough (1957: 254). Similar point can also be found in Rawls (1999a: 115).

¹⁹ See, for example, Scanlon (1998: 170-171), Scanlon (2008: 2-4) and Rawls (1999a: 98-99).

Chapter 2 Why Contractarianism?

contractarian principles has a virtue, for it represents a voluntary scheme which is rational to free and equal persons.

No society can, of course, be a scheme of cooperation which men enter voluntarily in a literal sense; ... Yet a society satisfying the principles of justice as fairness comes as close as a society can to being a voluntary scheme, for it meets the principles which free and equal persons would assent to under circumstances that are fair. In this sense its members are autonomous and the obligation they recognise self-imposed.²⁰

Some contractarians even believe that this emphasis on the importance of acceptability is the central virtue in the long tradition of contractarianism.²¹ One of the examples is T. M. Scanlon, who believes that no normative principle can be imposed on people unless it can be demonstrated that it is acceptable to these people.

Fundamental economic and political institutions cannot be justified simply on the ground that those to whom they apply have consented to their authority. Nonetheless, in order to be justifiable, institutions must give individuals the power to shape many of their particular obligations through the choice they make. The fact that an individual has chosen a certain outcome, or could have avoided it by choosing differently, is often an important reason why that outcome is legitimate.²²

[A]ccording to my version of contractualism, one needs to make claims about the reasons that individuals have for accepting or rejecting certain principles as standards of conduct... This

²⁰ Rawls (1999a: 12).

²¹ Freeman (2007b: 6).

²² Scanlon (2003a: 5).

Chapter 2 Why Contractarianism?

emphasis on what principles others have reason to accept (or reject), and on our reasons for caring about this, led me to call my view ‘contractualist.’²³

2.2.3 The virtue of respecting plurality

Thirdly, the contractarian justification of principles does not need to assume that any substantial moral consensus exists in the society. Modern society is highly pluralistic. People usually have diverse, or even conflicting, perspectives on how political society should be arranged. On the view of some people, such as Rawls, this situation is an inevitable consequence of liberal democracy, which allows people to develop their own viewpoints freely.²⁴ If a political principle is based on some substantial, controversial values, then it may have to assume that everyone in the society takes these substantial values as fundamentally important, which is obviously an unrealistic assumption in most modern societies. As Rawls indicates, a unified moral consensus can only be achieved by the oppressive use of state power: ‘In the society of the Middle Ages, more or less united in affirming the Catholic faith, the Inquisition was not an accident; its suppression of heresy was needed to preserve that shared religious belief’.²⁵

The strength of contractarianism is that it can justify political principles without making such a problematic assumption. The contractarian justification usually relies on some less controversial, commonly accepted ground, for example, rationality. It does not mean that contractarian justification is neutral, for some contractarians also acknowledge that their contractarianism clearly has some normative assumptions.²⁶

²³ Scanlon (2004: 124-125). Rawls emphasizes on the idea of justifiability can also be found in Rawls (1980: 305).

²⁴ Rawls (1993: 4).

²⁵ Rawls (1993: 37).

²⁶ Rawls (2001: 153).

Indeed, it is also impossible to have a theory which is completely normatively neutral.²⁷ Despite their partiality, contractarians at least are highly aware of the plurality of modern societies and try their best to avoid controversial justifications which are not publicly justified. The ideal picture of contractarianism is always that, even if people have widely divergent perspectives about political arrangements the political principle is still justified by some grounds which are commonly shared among these people.

Gauthier is one of the contractarians who emphasises this virtue of contractarianism. Gauthier's contractarian principles are justified by a weak assumption of instrumental rationality which is compatible with many interests and preferences. Although the society is highly pluralistic and conflicts are inevitable, the contractarian political principles can still be publicly justified to people. No matter what interests or preferences people have, all of them share the rationality of satisfying these interests or preferences effectively. Thus people will agree upon Gauthier's contractarian principles given that following these principles can lead to a mutually beneficial outcome.²⁸ These principles can still be the focus of social agreement even though an antecedent substantial consensus is absent. This awareness of plurality is commonly seen in the writings of other contractarians as well.²⁹ No contractarian would simply assume that certain substantial moral doctrines are widely accepted, or propose that the diversity of modern societies should be replaced by a substantial moral consensus. Since the fact of pluralism is hard to deny, this realistic understanding of society can be seen as one of the virtues of contractarianism.

²⁷ Milo (1995: 197).

²⁸ Gauthier (1986: 102-103).

²⁹ See Rawls (1993: 30-37). Similarly, although Scanlon presupposes that all people have an inclination to be reasonable, it does not mean that all of them are holding the same moral doctrine. They can still develop many different moral doctrines, insofar as these doctrines are compatible with the demand of reasonableness. See Scanlon (1998: 338-340).

By having a better understanding of the virtues of contractarianism, we can understand why so many political theorists are interested in this approach. Moreover, one can also see the distinctiveness of contractarianism. Contractarianism is not a wholly freestanding methodology which is compatible with any political theories. It justifies principles without relying on controversial metaphysical backgrounds, it respects the rational will of the individual, and it provides a public ground of justification in a pluralistic society. These three virtues are the strengths of contractarianism and can also explain why this methodology has been revived in contemporary political philosophy.

2.3 The ‘standard critique’: why am I bound by a hypothetical contract?

The virtues of contractarianism have been identified, but the weaknesses of this methodology should not be ignored. Some critics argue that the source of normativity of hypothetical contract is doubtful. The obligation which would be accepted by a hypothetical person cannot constitute a reason for action for a real person. This critique was first raised by Ronald Dworkin, and soon became the ‘standard critique’ of contractarianism.³⁰ I take this critique as pointing out the most serious weakness of the contractarianism that I am discussing, because if this critique is sound, then the hypothetical contract merely specifies political principles with no normative force. Hence, in this section, I will concentrate on this standard critique and discuss how contractarianism might in fact be immune from this critique.

³⁰ Stark (2000: 313-314).

2.3.1 The necessity of idealisation

Before I discuss this critique, it would be helpful to discuss the role of idealization in this theory. Contractarianism is sometimes misunderstood as relating to a hypothetical question about actual reactions, that is, whether the principles would be agreed to by people who live in the real world if they were asked. Indeed, if contractarians are really concerned with this issue, then what they would need to do is to conduct an empirical survey of actual people. However, what contractarians care about is actually the question of whether their principles would be agreed to by hypothetical contractors. It is rather a *doubly* hypothetical question, that is, a *hypothetical* question about a *hypothetical* reaction. Contractarians are concerned with neither what actual agreement would be reached by actual people, nor whether their hypothetical contract would be agreed to by actual people.³¹ This question is then a theoretical rather than an empirical question.

Given that contractarians are concerned with a doubly hypothetical question, the worry of critics should be clear. What is the point of caring about this doubly hypothetical question? Why is the perspective of hypothetical, idealised people relevant to the perspective of actual people? There are two reasons which can justify the necessity of studying this doubly hypothetical question. The first reason is one of practical limitation and the second reason is one of normative significance. First, given that the opinions of real people are so diverse, it is impossible to reach a unanimous agreement with everyone in a real society. There will always be some dissenters. Even if such a unanimous agreement is possible, it may be so vague that it cannot provide any substantial guideline for policy-making. If contractarians want to get some substantial principles through the device of hypothetical contract, then the

³¹ This can be seen in Gauthier's double denial. See Gauthier (1991a: 101).

idealization of this contract is necessary. As Rawls acknowledges, ‘without [idealisation] we would not be able to work out any definite theory of justice at all’.³²

Secondly, even if a unanimous, substantial actual agreement is possible, this agreement might be affected by some irrelevant factors. Suppose that a political principle is able to be unanimously agreed by everyone in a society. Nevertheless, some of them might agree with this principle because they have been coerced, because they lack information, or because of their ideological beliefs.³³ In this case, this principle does not have too much normative significance and we intuitively think that people do not have an obligation to comply with this political principle. Hence, an idealised hypothetical contract is necessary in order to show what principles people would agree with in a situation viewed as morally untainted.

These two reasons can explain why contractarians insist on justifying principles by a purified, idealised hypothetical contract, for the content of the agreement can become more determinate and irrelevant factors can be excluded. Without idealisation, the contractarian principles may either be too vague, or be corrupted by enforcement and fraud.

2.3.2 The standard critique

Idealisation prevents contractarianism from becoming desperately complicated or a product of force, but it might be gained at the expense of vacuousness. Contractarianism becomes vulnerable to the critique of being irrelevant to the real political world. Since some factors affecting people’s reasoning are excluded in the

³² Rawls (1999a: 121).

³³ Gaus (2011: 26).

idealised choice situation, the decision of a hypothetical person in the idealised choice situation is surely different from the decision of an actual person in the real world. Because of this gap, actual people might not agree with the hypothetical agreement even if they were asked; given that I have not agreed with this hypothetical agreement in my life, and I would not agree with it if I were asked, why should I be bound by this hypothetical agreement?

This standard critique has often been mentioned in the literature.³⁴ It is usually represented by Dworkin's critique of Rawls' idea of original position, which argues that a hypothetical consent cannot generate any obligation to real people.

I might concede that I would have agreed for that reason, and then add that I am lucky that you did not raise the point. The fact that I would have agreed if you had insisted neither adds nor suggests any argument why I should agree now. The point is not that it would have been unfair of you to insist on your proposal as a condition of playing; indeed, it would not have been. If you had held out for your proposal, and I had agreed, I could not say that my agreement was in any way nullified or called into question because of duress. But if I had not in fact agreed, the fact that I would have in itself means nothing.³⁵

That is why Dworkin famously suggested that 'A hypothetical contract is not simply a pale from an actual contract; it is no contract at all'.³⁶ Hypothetical contract seems to assume that *you* can be bound by an agreement that *others*, different from you, would have made. This seems unconvincing because the relationship between the hypothetical contract and actual people is unclear. Thus some critics mischievously

³⁴ Similar view can also be found in Dworkin (1973), Nagel (1973), Fiskin (1990), Brudney (1991), Hampton (1997: 65-66), Hampton (1993b), Simmons (1993b: 220-221), Brown (1988: 494-495), Wolff (2006: 48-50), Harman (1977: 111), Atiyah (1979: 143-146) and Kelly (2004: 230).

³⁵ Dworkin (1973: 18-19).

³⁶ Dworkin (1973: 18).

say that hypothetical contract is merely ‘good clean intellectual fun’.³⁷ Other critics, such as Dworkin and Pettit, are more sympathetic to this methodology, but still claim that contractarianism is incomplete unless it is supplemented with an additional moral theory. Dworkin argues that what is behind contractarianism should be a theory of natural rights. Since everyone has a natural right to equal concern, we are obligated to give each person equal respect.³⁸ The hypothetical contract is a medium for deriving substantial principles from abstract natural right. Pettit believes that contractarianism threatens to collapse to a kind of consequentialism.³⁹ When contractors reject a certain principle, they usually reject on grounds of certain impersonal values, such as fairness and happiness.⁴⁰ So what ultimately matters should be these impersonal values but not the rejection of hypothetical contractors. A political principle is unjustified not because it would be rejected in a hypothetical contract, but rather is because it fails to promote certain impersonal values.

2.3.3 The orthodox reply and its inadequacy

Theorists who believe that contractarianism is a self-sufficient, complete theory intend to clarify that contractarianism is not a theory of obligation and authority, but rather a theory of content. It does not try to define what normative obligation people actually have. Rather, it only aims at proposing an appropriate decision procedure for political philosophy. A hypothetical contract merely demonstrates an ideal situation that ideal contractual parties would reason properly and would agree on several political principles. This contract is a heuristic device which clarifies the content of political principle. Through the hypothetical contract, people can know how political

³⁷ Baier (1994), quoted in Hampton (1993b: 2).

³⁸ Dworkin (1973: 50-52).

³⁹ Although Pettit is concerned with moral contractarianism and argues that contractarianism cannot provide a satisfactory account of moral wrongness, his objection can also be applied to political contractarianism. For Pettit’s critique, see Pettit (2006).

⁴⁰ Pettit (2006: 79).

principles are ‘constructed’ through certain decision-making processes. But this philosophical project can be conducted without requiring a theory of obligation and authority. Contractarians can acknowledge that the hypothetical contract itself cannot create any political obligation. Whether actual people are obligated to these political principles and whether they would accept this obligation are beyond the concern of contractarians. This strategy is an orthodox answer given by theorists who try to defend contractarianism.⁴¹ One of the examples is Jean Hampton:

If we understand the structure and the role of the contractarian device in our moral thinking, the contract idea isn’t in any sense foundational, or even necessary, for effective moral reasoning. It is merely a test that is heuristically valuable for the moral agent in virtue of the fact that it is informed by ideas that they are the real source of moral reasoning.⁴²

It is true that the hypothetical contract *itself* is inadequate to generate political obligation, for it is hard to deny that there is a gap between hypothetical contractors and actual people. However, it seems that most of the contractarians do not think their theories are *merely* theories of content. Although many contractarians are concerned with the problem of content,⁴³ they also believe that the problem of obligation and authority can be solved by their theories. For example, Rawls believes

⁴¹ The idea that a hypothetical contract is only a metaphor can be seen in Morris (1988). Also, someone might suggest that Rawls should be understood as proposing a theory of content. For example, Rawls argues that, in his contractarian theory, “the principle of political justice (content) may be represented as the outcome of a certain procedure of construction”. (Rawls 1993: 89-90) Through the hypothetical contract, the content of justice and the order of ethical principles can be clarified. Rawls believes that this is a virtue of contractarianism and contrasts it with intuitionism, which argues that there is an ultimate plurality of ethical principles and no procedure can weigh these principles when they conflict.

⁴² Hampton (1993b: 30).

⁴³ Rawls (1996:89-90), Gauthier (1986: 9) and Barry (1989).

that every individual has a natural duty of justice, which ‘do[es] not presuppose an act of consent, express or tacit, or indeed any voluntary act, in order to apply’.⁴⁴

From the standpoint of justice as fairness, a fundamental natural duty is the duty of justice. This duty requires us to support and to comply with just institution that exist and apply to us. It also constrains us to further just arrangements not yet established, at least when this can be done without too much cost to ourselves. Thus if the basic structure of society is just, or as just as it is reasonable to expect in the circumstances, everyone has a natural duty to do his part in the existing scheme. Each is bound to these institutions independent of his voluntary acts, performative or otherwise.⁴⁵

Rawls is not alone at this point. Gauthier and Nagel also believe that a hypothetical contract can justify enforced compliance to political principles.⁴⁶ In general, these contractarians do not treat the hypothetical contract as, following Hampton, ‘just a way of reasoning that allows us to work out conceptual answers to moral problems’ and has nothing to do with whether the actual people will accept this way.⁴⁷ Rather, they believe that this hypothetical consent can be used as a substitute for actual consent. So the contract which is justified by this hypothetical consent has normativity, regardless of what responses actual persons have.

This shows that even if a gap between hypothetical contractors and actual people exists, contractarians still think that coercive obligations to particular principles can be justified by a hypothetical contract and we, as actual people, are truly bound by this hypothetical contract. But where does the normativity of hypothetical contract

⁴⁴ Rawls (1999a: 99).

⁴⁵ Rawls (1999a: 99).

⁴⁶ See Nagel (1991: 36-37). Similar point can also be found in Gauthier’s writings, though he is talking about obligation to particular moral principles. See Gauthier (1986: 117).

⁴⁷ Hampton (1993b: 14). Gauthier clearly disagrees with this point. See his Forward in Hampton (2007: xii).

come from? Indeed, some critics seem to think that this binding power is merely an illusion, since only actual consent can justify coercive obligations.⁴⁸ If a person consents only in a hypothetical situation but does not consent in an actual situation, this consent has no normative force for anyone. However, this account of coercive obligation is too narrow. The standard critique sets a very high standard for imposing political obligations: it assumes that the imposition of obligation must be voluntary. It is an unrealistic standard because, in fact, most people have never voluntarily expressed their consent to their political obligations. If this assumption were true, then most of the people in the world would have no political obligations to their states, which is absurd, or at least inconsistent with any non-anarchist stance.⁴⁹ Moreover, sometimes even when a person actually consents to something, it is still unjustified to impose obligations on him. Not every actual consent creates an obligation.⁵⁰ For example, if people are misinformed or threatened, their actual consent should not have any normative force. Therefore, actual consent is neither a necessary condition nor a sufficient condition for coercive obligation. If actual consent is irrelevant, then it means that even without actual consent, a hypothetical contract can still have normative force. But what are the conditions?

2.4 The normativity of hypothetical contract

The standard critique is correct in that hypothetical consent should not be confused with actual consent, yet it is too early to claim that normativity can only come from actual consent. So when does a hypothetical contract establish normativity? Some

⁴⁸ Stark (2000: 325-327). Clearly Dworkin does not hold this view since he believes that the obligations to hypothetical contracts come from natural rights, and the existence of natural rights is not affected by actual consent.

⁴⁹ The problem of voluntarism has also been discussed in other works, such as Simmons (1993a) and Wolff (1998).

⁵⁰ Consent has normative significance only when it satisfies certain conditions. Not all voluntary consent can reflect the value of choice. See Scanlon (2001: 268).

theorists argue that hypothetical contract is justified when it is supported by ‘good reasons’.⁵¹ If people have good reasons to accept the premises of a hypothetical contract, then they should also accept the hypothetical contract. But what can be counted as good reasons? This account is too vague and we must go further to discuss the source of normativity of hypothetical contract. In this section, I will show that a hypothetical contract has normative force only when it can model practical reason correctly.

2.4.1 The normativity of practical reason

The source of normativity of the hypothetical contract can be made explicit by discussing the role that practical reason plays in contractarian theory. Here practical reason means a general human capacity for deliberating upon practical issues. Unlike theoretical reason, which determines what one ought to believe, practical reason is concerned with what one ought to do (which is why it is ‘practical’). It represents a capacity for thinking through a distinctive process. When people deliberate about actions and choices, they think about themselves and their situations through this process, and eventually come up with certain actions and choices. This is why practical reason also has a crucial role in explaining human motivations and behaviours. An action or choice can be justified insofar as it can be accounted for by practical reason. In short, practical reason determines the actions and choices of people.

Practical reason, as a process of thought, is not wholly arbitrary, for it is governed by certain antecedent normative rules. These constraints are unchosen and bind the

⁵¹ Stark (2000: 333-334) and Gaus (1990: 328).

practical thinking of human beings.⁵² One example of these rules is the syllogism mentioned by Aristotle. Given that people have certain beliefs, they are committed to certain actions or choices. Given that I want to take the train at 4pm, and if I do not leave home now I cannot reach the train station before 4pm, I should leave home now. In the following chapters, I will discuss these rules, which fall into two conceptions of practical reason: rationality and reasonableness. But at this stage I will only define practical reason as a capacity for deliberating upon practical issues in terms of certain rules. 'Practical reasoning is a vehicle for solving casuistic problem of what to do in a particular situation, given that certain...principles have been designated or established as personal commitments to be acted upon by an agent'.⁵³

The normativity of practical reason is independent of and more fundamental than other rules that bind people. There are many rules in the world. For instances, in the rules of football, we cannot play football with our hands. In the rules of chess, the King cannot be moved two squares. However, we can reflect upon our obligation to these rules. We do not necessarily have to follow them and can escape from them whenever we want. But we cannot avoid deliberating in terms of rationality or reasonableness, since these rules are actually the principles of *logic* of practical deliberation. These rules determine our way of reflection. Hence there is no sense at all in reflecting upon whether we should deliberate in terms of rationality and reasonableness, since reflection itself has to follow these rules. Asking about the obligation to these rules is merely begging the question.⁵⁴ Whenever we exercise practical reason, we necessarily commit ourselves to complying with these rules.

⁵² Similar point can be seen in Raz (1979b: 138-142).

⁵³ Becker and Becker (2001: 1355).

⁵⁴ Gibbard (1985: 37-39). Although Gibbard then suggests that the question of whether rationality is justified might be at the end able to be answered in the 'normative' way, I am not sure how this could not be circular since, in his proposal, morality is defined as a matter of how certain specific moral sentiments can be rational.

Commitment to such rules is inevitable in practical reasoning. Hence, the normativity of practical reason is independent and people are involuntarily bound by them.⁵⁵

Also, the normativity of practical reason is fundamental since the normativity of other rules is determined by practical reason. We are committed to complying with different rules in our lives, but whether these obligations are justified is determined by practical reason. For example, a person promises to follow a rule in a game. But if she then discovers that this rule is irrational/unreasonable, she will think that this promise is unjustified and does not have any binding power. In light of this, one can see that practical reason is the sources of normativity of other rules. Our commitments to complying with other rules are justified insofar as they are justified in the reasoning governed by rationality and reasonableness. Hence, one cannot reject the conclusions of practical reason because of other commitments in one's life.

One might doubt the involuntary normative force of practical reason, since one could imagine an insane person who ignores these rules. Taking the train example again, an insane person could want to take the train at 4pm, and know that he must leave home now in order to catch the train, but still refuse to leave home. This is biologically possible, yet it is also obvious that these cases are extremely rare. In most circumstances, people deliberate under the guidance of rationality and reasonableness. These rules are the normal pattern of human behaviour. It is surely possible that people not strictly follow this pattern, yet this is only because of the influence of some irrelevant factors, such as laziness, coercion and misinformation. If people could deliberate without being affected by these irrelevant factors, then they would generally follow the guidance of rationality and reasonableness. We simply think that the behaviour of an insane person does not make sense if this

⁵⁵ Railton (1997: 64).

person did this even after thorough deliberation. In short, although it is biologically possible to violate the rules that govern practical reason, it is impossible to do so intelligibly. Our practical reasoning is inevitably regulated by these normative rules.

2.4.2 Hypothetical contract and practical reason

Practical reason explains where the normativity of hypothetical contract comes from. The basic strategy of contractarianism is to explain the normativity of certain political principles in terms of practical reason, and the device is hypothetical contract.⁵⁶ The rules that govern practical reason can be applied to different questions, such as the question of political principles. This is the function of hypothetical contractors. They are defined as ideal agents who are placed in carefully circumscribed conditions and deliberate about certain questions without being affected by irrelevant factors. Their thought and judgment reveals the deliberative results of actual people when they think strictly in accordance with the rules that govern practical reason. In short, the hypothetical contractors model the practical reasoning of actual people. Contractarians then discuss what political principles would be chosen by these hypothetical contractors. Through discussing this scenario, contractarians show what will happen if the rules that govern practical reason are applied to the question of political principle. The political principles that are chosen by the hypothetical contractors represent the conclusion of practical reasoning when actual people deliberate upon the question of political principles correctly.

⁵⁶ Actually, this point is partly discussed by Gauthier in 'Why Contractarianism?' He proposes that the hypothetical contract is a 'rational reconstruction' and the reason why the political principles derived from it have normative force is because they are mutually, rationally agreed principles. See Gauthier (1991a: 105). Nevertheless, Gauthier still fails to explain where the normative force of hypothetical contract comes from, so his article still cannot answer Dworkin's 'standard indictment' well. This section is an attempt to push Gauthier's argument further.

An example of this way of modelling practical reason is the hypothetical contractors in Gauthier's contractarian theory. In this theory, hypothetical contractors are rational, fully-informed utility-maximizers. The purpose of defining the hypothetical contractors in this way is to show the pure practical reasoning of a rational person. In the real world, actual people might fail to deliberate fully rationally. Their practical reasoning might be influenced by some irrelevant factors, such as deception and short-sightedness. The rational utility-maximizers show what would happen when actual people deliberate in accordance with the rules of rationality. The choices of these hypothetical contractors reveal what political principles are rationally justified.

In light of the rationale behind contractarians' definitions, it is also clear why actual people should accept the choice of hypothetical contractors. Suppose we assume that the hypothetical contractors can perfectly model the practical reasoning of actual people. These hypothetical contractors deliberate in accordance with the rules that regulate the practical reasoning of actual people as well. Since the practical reasoning of hypothetical contractors and actual people is governed by the same normative rules, the result of their practical reasoning should be the same. If hypothetical contractors agree with certain political principles, then these political principles should be justified to actual people as well. These principles are principles that actual people should accept insofar as they can deliberate correctly. The function of hypothetical contractors is to help actual people to clarify their practical reasoning and uncover the obligations that actual people have reason to accept. As Rawls points out, his original position is indeed a kind of 'thought experiment for the purpose of public- and self-clarification'.⁵⁷ Provided that hypothetical contractors can model our practical reasoning correctly, we have no reason to reject the choices of hypothetical contractors; *contractors' choices are actually our choices*.

⁵⁷ Rawls (2001: 17).

However, one should not misunderstand that a hypothetical contract would be agreed by actual people if they were asked. In the real world, actual people might be affected by laziness, coercion, prejudice or misinformation, and hence might choose differently. As Gaus writes, ‘actual people do not always act on their sound and overriding reasons’.⁵⁸ Despite the weaknesses of actual people, the normativity of the hypothetical contract is unaffected, since its normativity is based on practical reason, not on the possibility of being actually agreed to. Why actual people make different decisions is simply because they are influenced by factors which are irrelevant to practical reasoning. Their practical thinking is still regulated by the normative rules that govern practical reason. If actual people were aware of the existence of these factors, they would also prefer excluding them from practical reasoning. Hence the choices of hypothetical contractors are justified to actual people regardless of their actual responses.

Therefore, the source of normativity of the hypothetical contract is made explicit after we understand the relationship between hypothetical contract and practical reason. *A hypothetical contract has normative force when the hypothetical contractors can model the practical reasoning of actual people adequately.* Our practical reasoning is governed by certain normative rules, and it is impossible intelligibly to deny the normativity of these rules. The hypothetical contractors are defined for the sake of showing the process of practical reasoning when these rules are strictly followed. The political principles chosen by hypothetical contractors represent the conclusion of proper practical reasoning. The normativity of these principles is explained by practical reason.⁵⁹ Hence, actual people have good reasons

⁵⁸ Gaus (2011: 28).

⁵⁹ Someone might disagree with this account and argue that the source of normativity of hypothetical contract is the conceptions of person and society, but not the conception of practical reason. For example, Rawls presupposes conceptions of person as free and equal citizen and society as a fair system of social cooperation. Then his contractarian theory is constructed on the ground of these political ideals. However, if we understand Rawls’s conceptions of person and society properly, then

to accept these political principles, since the reasoning of hypothetical contractors is actually *their* reasoning. The gap between hypothetical contractors and actual people is bridged by practical reason.

2.4.3 The problem of the standard critique

Once the source of normativity of hypothetical contract becomes clear, we should be able to see the problem with the standard critique. The critique is based on the wrong question: ‘why should I accept the hypothetical contract?’ The reason for asking this question is understandable, since the hypothetical contract seems to argue that one can be bound by the consent of fictional characters that are *irrelevant* to one’s concern. Hypothetical contractors legislate for actual people and create some obligations that actual people would not agree with. It seems that the will of actual people is ignored and thus some theorists question why we should be bound by this contract.

However, the theorists who offer this critique misunderstand the relationship between hypothetical contractors and actual people. As I discussed in the last section,

we can see that these conceptions are not independent sources of normativity. Rather, these conceptions merely represent the context in which people apply the rules that govern practical reason, which are the true source of normativity. As Rawls indicates, these conceptions “characterize the agents who reason and they specify the context for the problems and questions to which principles of practical reason apply”. (Rawls 1993: 107) We do not know the implications of rationality and reasonableness unless we apply these principles to certain problems. For example, if we only know that a person takes a flight to Paris, then we cannot evaluate whether or not this act is rational. But if we further know his problem and his context (He has a meeting at Paris. How can he arrive at there as soon as possible?), then we can discuss how rational his act is. The function of the conceptions of person and society is to provide a background for the principles of rationality and reasonableness to be applied. People imagine themselves as free and equal citizens who live in a fair system of social cooperation, and they know that they have to choose principles of justice. Then they deliberate in accordance with principles of rationality and reasonableness. In short, the conceptions of person and society define the practical questions, while the rules that govern practical reason define the ways to deliberate upon these questions. I am indebted to Jonathan Quong for the discussion about this point.

provided that the hypothetical contractors are properly defined, hypothetical contractors and actual people share the same methods of practical reasoning. The point of view of hypothetical contractors is *my* point of view. Actual people should accept the hypothetical contract, since they also deliberate in accordance with the same normative rules. They can only recognise the results of proper practical reasoning, but not reject them. So the hypothetical contract does not *create* any obligations; it merely *discovers* the obligations that actual people already have. The hypothetical contract is not a story that a fictional, irrelevant character legislates for you. Rather, it is a thought experiment that helps you to clarify what you should legislate for yourself in the political realm. Hence, the question of why we should accept the hypothetical contract is misleading, since the hypothetical contract is not an external constraint that is imposed by some authorities outside of us. Rather, it is an internal constraint that we are already committed to but we have not realised. The hypothetical contract simply shows us what we *should* think and, indeed, *do* think.

This understanding of the hypothetical contract can also be seen in the writings of some contractarians. For example, Rawls believes the original position is important in clarifying the implications of our conception of the person, which is characterized by rationality and reasonableness. '[The hypothetical contract] helps us to work out what we now think'.⁶⁰ Hence, the principles derived from the original position have normative force to us because these principles are justified by the conception of the person that 'we do in fact accept'.⁶¹ The aim of the original position is not to create any obligations, but 'to uncover a public basis for a political conception of justice'.⁶² It merely shows us what political principles we should accept, given that we have a particular conception of the person.

⁶⁰ Rawls (1993: 26).

⁶¹ Rawls (1999a: 514).

⁶² Rawls (2001: 81).

Chapter 2 Why Contractarianism?

The context of the problem guides us in removing vagueness and ambiguity in the conception of the person, and tells us how precise we need to be...Thus the structure defined by the original position may enable us to crystallize our otherwise amorphous notion of the person and to identify with sufficient sharpness the appropriate characterization of free and equal moral personality.⁶³

In light of this relationship between practical reason and hypothetical contract, we can also see why those theorists who claim that contractarianism is incomplete wrongly understand the nature of contractarianism. Dworkin suggests that hypothetical contract indeed assumes a self-evident theory of natural rights. However, he misunderstands the real assumption of contractarians. In most of the contractarian theories, individual rights are not taken for granted; rather, they are the product of people's agreement and are justified by their conceptions of practical reason.⁶⁴ In the discussion above, we can see that the source of normativity of the hypothetical contract is practical reason, not natural rights. If contractarians take natural rights as the source of normativity, then they have to worry about explaining the relationship between the world and these natural rights. But if they take practical reason as the source, then this worry can be avoided because the rules that govern practical reason are merely general patterns of thought that are induced by empirical studies of human behaviour.

Pettit argues that contractarianism must presuppose a consequentialist theory of promoting certain impersonal values. What matters is not the fact that the contractors would reject certain principles, but rather the consequences that certain impersonal values would be promoted by following these principles. These consequences are the

⁶³ Rawls (1980: 357).

⁶⁴ The refutation from Rawls to Dworkin's definition is a good example which shows the inadequacy of treating hypothetical contract as a right-based theory. Rawls clarifies that his position should be seen as 'conception-based' or 'ideal-based' insofar as it works from ideal conceptions of both person and society. 'Right, duties and goals are but elements of such idealized conceptions' (Rawls 1985: 400n. 19). Similar views can be found in Gauthier (1991a: 98) and Nagel (1970: 18).

reasons for contractors to reject these principles. However, contractarians resist this consequentialist assumption. They argue that the rejection of rational or reasonable contractors is the fundamental ground of rejecting a political principle. The consequentialist considerations of promoting impersonal values would be considered by those rational or reasonable contractors, but these would not be the only considerations. There are some other non-consequentialist considerations which cannot be reduced to consequentialist considerations, such as reasons to keep one's promise.⁶⁵ Rational or reasonable contractors would deliberate in accordance with the rules that govern practical reason, weigh these considerations and choose political principles. There is no an consequentialist theory behind contractarianism. Rather, contractarianism is a self-sufficient theory which reveals our practical reasoning and takes consequentialist considerations into account.

One distinction may help us to see how Dworkin and Pettit misunderstand the function of hypothetical contract. In Plato's *Euthyphro* Socrates asks whether something is holy because the gods love it, or whether the gods love it because it is holy. Similarly, we can ask whether a political principle is justified because it would be chosen in the hypothetical contract, or whether it would be chosen in a hypothetical contract because it is justified. For the former case, the hypothetical contract is definitional of what it is to be justified. For the latter case, the hypothetical contract merely identifies principles that are justified by other independent grounds. Dworkin and Pettit think that the latter account for hypothetical contract is correct. Hypothetical contract merely identifies principles that protect natural rights or promote impersonal values. However, most of the contractarians think that hypothetical contract should be definitional. For example, Rawls proposes the idea of procedural justice and argues that the original position is

⁶⁵ Scanlon (2000:243-245)

the correct procedure of defining principles of justice.⁶⁶ Apart from this procedure there are no independent criteria of justifying political principles. Gauthier and Scanlon also define principles of right in terms of what would be agreeable to rational or reasonable parties.⁶⁷ Hence, the contractarian principles are defined by hypothetical contract which reveals proper practical reasoning and are not justified by other independent normative grounds.

In conclusion, I do not intend to disagree with the standard critique of hypothetical contract. The standard critique is correct that, because of idealisation, hypothetical consent cannot be taken for actual consent. However, as long as hypothetical contractors can model the practical reasoning of actual people correctly, the normativity of the hypothetical contract is justified, regardless of whether or not a person actually consents to it. The political principles specified by the hypothetical contract represent the requirements of practical reason. In short, the function of the hypothetical contract is not to *create* political obligations, but to *discover* political obligations that are justified by practical reason.

2.5 Two conditions: the condition of generality and the condition of priority

In the last section, I showed that the normativity of hypothetical contract depends on whether the definition of hypothetical contractors can correctly model practical reasoning. Therefore, a contractarian theory must be based on a conception of

⁶⁶ Rawls (1980). Here my example of Plato's *Euthyphro* is borrowed from Kukathas and Pettit (1990: 28-29), but my understanding about Rawls is different from theirs. Kukathas and Pettit believe that Rawls's usage of hypothetical contract is the latter case, that is, he presupposes an independent criterion of fairness and the original position is only used to identify principles which meet this criterion. However, this interpretation is incompatible with Rawls's idea of procedural justice. To Rawls, original position *is* the fair procedure, and there is no a criterion of fairness independent of the original position. Thus the role of the original position in Rawls's theory should be definitional, which is the way that hypothetical contract is used in the former case.

⁶⁷ Gauthier (1986) and Scanlon (1982)

practical reason, an assumption about the rules that govern the practical reasoning of actual people. However, there are so many ways of modelling practical reasoning and how should a contractarian choose? As I mentioned in Chapter 1, political principles must satisfy two conditions: the condition of generality and the condition of priority. If a contractarian theory aims at justifying political principles, then its conception of practical reason must help it satisfy these two conditions. In this section, I will discuss how a conception of practical reason can perform this function.

2.5.1 The condition of generality

As I mentioned in Chapter 1, the condition of generality is a requirement of *scope*, to the effect that the political principle should be widely accepted by people to which it is going to apply. In order to satisfy this condition, a contractarian theory should be based on an account of practical reason which is widely accepted as well.⁶⁸ As Andreas Eshete points out, the normative force of a hypothetical contract will be weakened if the design of this hypothetical contract is not based on some general descriptions.

If the contractarian program is to succeed, it is essential that the principles adopted on the basis of the primitive facts of justice be acceptable to men in any specific situation within the range of general facts...Hence, contractarian theories are undermined when it is shown that they

⁶⁸ I specify the scope of the application of contractarian theories, for not all of the contractarians aim at justifying political obligations to everyone. For example, the later Rawls aims only at justifying our obligations to principles of justice in a liberal democratic society. We are obligated to principles of justice because we identify ourselves as free and equal citizens of a liberal democratic society. Therefore we accept the choice of hypothetical contractor in the original position. But for non-liberal-democratic societies, it is another story. However, Rawls still has to prove that, to those citizens of liberal democratic society, the conception of the person as free and equal citizen is generally acceptable.

Chapter 2 Why Contractarianism?

favour a particular model of society, a partisan conception of the good, or a special interpretation of human psychology.⁶⁹

This is why contractarians usually rely on rules that govern practical reason, but not other rules in our lives, to justify the normativity of the hypothetical contract. The normative force of many rules in our lives is contingent. These rules are contingent in that whether one is bound by them depends on one's contingent circumstances. We acquire these rules in various ways. Some, such as our profession or religion, are chosen by us. Others we are born into, such as being someone's son or being a member of a certain culture. However, all of these rules merely bind us contingently. If we were born in another place, then we would follow the rules of another family. Similarly, if we grew up in another context, then we might choose another job. Unlike these contingent rules, rationality and reasonableness, which are the rules that govern the process of thought, are unconditional. People can only think within the constraints of these rules. No matter which family one belongs to, which choices one makes, the normative force of rationality and reasonableness are unaffected, for these rules bind people irrespective of their personal context.

However, there are various interpretations of these rules, and not all interpretations provide a suitable basis for contractarianism. Some are substantial conceptions such that only a particular group of people really think in accordance with these rules. For example, the Thomistic conception of rationality assumes that human beings share an ultimate end in life, that is, to pursue eternal happiness in Heaven. All rational human behaviours should serve this ultimate purpose. Although this interpretation is called rationality, it is not generally acceptable since many non-Catholics do not think that this conception correctly describes their pattern of reasoning. Unlike the Thomistic conception, there are some formal conceptions of rationality that are more acceptable

⁶⁹ Eshete (1974: 40).

to many people, from a wide range of backgrounds. For instance, most people believe that their patterns of reasoning are guided by an instrumental conception of rationality. This conception of rationality is compatible with a wide range of ends and it merely describes that rational people will pursue their ends effectively. In order to make sure that the normativity of contractarian principles will not be affected by any contingent factors, contractarians should adopt the latter kind of conception of practical reason rather than the former.

Hence the conception of practical reason assumed in a contractarian theory should represent a common way of practical thinking shared among persons holding different substantive views. This is the condition of generality. *Satisfying the condition of generality means that the conception of practical reason is generally acceptable to the actual people who are bound by the political principles.* Only when the rules that guide practical reasoning of the hypothetical contractors are general rules, can actual people generally be identified themselves as the hypothetical contractors. The choice of the hypothetical person is then not a completely strange, irrelevant choice to actual people.

2.5.2 The condition of priority

Apart from the condition of generality, the condition of priority is also necessary for a hypothetical contract to justify political principles. As I mentioned in Chapter 1, this condition is a requirement of *status*, the status of the reason for following political principles in a person's practical reasoning. To those people who accept the political principle, this reason should have lexical priority (or very strongly weighted priority) over other considerations. If there are some acts that are demanded by political principles, we cannot escape them because of our other considerations. This feature of political principle is widely accepted among contractarians. When contractarians justify our obligation to certain political principles by hypothetical

contract, they do not take this obligation as simply one which is meant to weight up against other competing considerations.⁷⁰

The key to justifying this deliberative priority is to rely on the ultimate normative authority of practical reason. It would be helpful to clarify the relationship between practical reason and the justifiability of other social norms. There are various social norms in people's lives, such as family norms, religious norms and cultural norms. Political principles are some of them. These social norms impose different obligations on these people. They then have to engage in practical reasoning and reflect on whether or not they should comply with these social norms. Suppose eventually their conclusion of practical reasoning is that political principles should have the highest priority. Then it is unjustified for them to violate political principles because of social norms. This is because, when these people decide whether they should comply with political principles or social norms in their practical reasoning, these social norms have already been taken into account. So they no longer constitute reasons for them to ignore this conclusion.

From this example we can see the difference between the rules that govern practical reason and other norms. They are rules of different natures. For the former kind of rules are rules that govern the *a priori* structure of practical reasoning and guide how we should think. For the latter kind of rules, they are merely considerations that we take into account when we engage in practical reasoning. Whether our obligations to the latter rules are justified or not depends on practical reasoning that is governed by the former rules. If the former rules do not justify obligations to the latter rules, then we have no reason to comply with the latter rules. In short, the relationship between

⁷⁰ See Rawls (1999a: 3-4, 263-264) and Scanlon (1998: 160-168).

the rules that govern practical reason and other norms is not coordinate. Rather, they stand in a relationship of superordinate and subordinate.

Hence, if a contractarian can demonstrate that a hypothetical contractor, who thinks in accordance with the rules that govern practical reason, will give political principles an overriding status, then actual people are unjustified in going against these political principles because of other norms. The condition of priority is fulfilled since the priority of political principles is the conclusion of proper practical reasoning which has taken obligations to other norms into account.

However, there is more than one rule governing practical reason. What if these rules conflict with each other? These rules represent different approaches to resolving conflicts of reason and establishing that which one should do. By adopting different approaches, the conclusion of practical reasoning might be different. For example, both rationality and reasonableness are rules that govern practical reason, and it is commonly argued that there is a sharp opposition between these two rules.⁷¹ Reasonableness requires people to take due account of the interests of others and to maintain a mutuality of consideration between themselves and others during practical reasoning. Rationality is a matter of means-end calculation that is concerned with the way of pursuing particular ends effectively. These two rules might justify opposite reasons for action since sometimes the most effective way to pursue certain ends involves violating the equitable relationship between oneself and others. While both of these rules are the fundamental constraints that govern practical reasoning, their normative forces are the same. People have fundamental reasons for action to pursue their ends effectively, but also have a fundamental reason for action to maintain the

⁷¹ Sibley (1953) and Held (1977).

equitable relationship as well. Neither of these reasons can necessarily override the other.

This implies that if a conception of practical reason fails to include all the rules that govern practical reason, then the priority of the political principles cannot be guaranteed. Suppose that a contractarian takes rationality as the whole of practical reason, defines hypothetical contractors as rational pursuers of ends, and eventually comes up with an agreement that specifies certain political principles. To actual people, they have an obligation to these principles because these principles are justified by rationality. But on the other hand reasonableness, another aspect of practical reason, justifies obligations that violate these political principles; so actual people will have two conflicting obligations. Rationality justifies obligations to follow political principles, whereas reasonableness justifies opposite obligations. Since these two obligations have the same weight in practical reason, actual people have no need to choose either of them.

In order to avoid this problem, contractarians should make sure that their conceptions of practical reason do not capture only a part of practical reasoning. They should attribute all the rules that govern practical reason to the modelled contracting parties, and show that all these rules justify their hypothetical contract. This is the condition of priority. *Satisfying the condition of priority means that all the rules that govern our practical reasoning are included in the conception of practical reason and shown to support the political principles.* Only when this condition is satisfied, can the deliberation of hypothetical contractors be seen as modelling the complete picture of practical reasoning of actual people. If not, the political principle may fail to possess an overriding status in people's practical reasoning.

2.5.3 Case study: John Harsanyi

The importance of these two conditions becomes clearer if we look at one contract model which fails to satisfy this condition. I will take John Harsanyi's contract model as a counter-example. Although Harsanyi is a rule-utilitarian,⁷² his theory also has contractarian elements, as he uses the idea of rational bargaining in a hypothetical contract to justify the utilitarian principle.⁷³ Harsanyi asks us to imagine that hypothetical individuals are behind a veil of ignorance, where they are unaware of their own social and economic positions and their own capacities and talents.⁷⁴ They know that they have equal probability to become anyone of the society's members, but now they have to choose a political principle for this society. In this idealised choice situation, the hypothetical persons become impartial sympathetic observers. They are sympathetic because they take the utility function of each participant into account. They are impartial because they are not biased in favour of any participant.⁷⁵ From the perspective of these impartial sympathetic observers, the political principle should be the principle of average rule-utilitarianism. However, why is the choice of these hypothetical contractors relevant to us? Harsanyi answered that actual people should agree with the choice of these hypothetical contractors because actual people, who are rational, should have a *moral preference*, which is a preference 'to judge the world from a moral, i.e., from an impersonal and impartial, point of view'.⁷⁶ Because of this preference, actual people will accept the principle of average rule-utilitarianism, and this is why the hypothetical contract story is not only irrelevant.

⁷² Harsanyi acknowledges this philosophical position in Harsanyi (1977).

⁷³ Harsanyi emphasizes that his model and Rawls's are similar contractarian models, see Harsanyi (1982: 47). The contractarian character of Harsanyi can also be seen in Harsanyi (1975) and Harsanyi (1958). For the interpretation of Harsanyi as contractarian, see Binmore (2008) and Boucher and Kelly (1994b: 22-23).

⁷⁴ Harsanyi (1982: 44-46).

⁷⁵ Harsanyi (1986: 49).

⁷⁶ Harsanyi's answer to this question is always ignored by commentators, and his answer can show his contractarian concern. See Harsanyi (1976: ix), Harsanyi (1975: 45-46) and Harsanyi (1982: 62).

Harsanyi's contract theory is based on a conception of rationality that views people as rational when they are guided by a moral preference. However, this conception of practical reason is unrealistic because the moral preference assumed by Harsanyi is rarely seen, even when people are rational. According to Harsanyi, because of this moral preference, we are willing to identify ourselves with the average utility-maximizer in the hypothetical contract. But, as an average utility-maximizer, one might have to make enormous sacrifices for the sake of maximizing the average utility of the whole society. These enormous sacrifices include giving up some basic liberties, such as liberty of conscience and freedom of thought. Therefore, having a moral preference and being an average utility-maximizer mean that one has to be unlimitedly altruistic to the extent that one is often willing to sacrifice one's fundamental interests. In light of our observations of human nature, this conception of practical reason is too altruistic to be generally acceptable to most human beings. Human beings usually have some fundamental interests that they must protect if they can, such as the interest in choosing and revising their ends of life.⁷⁷ No matter what interests they have, they still want these interests to be protected. Unless they are very altruistic, individuals are not willing to sacrifice these interests for others and are willing to be merely a means to maximize the average utility of the whole society. Thus, the conception of practical reason in Harsanyi's theory can hardly be seen as a general description of practical reason.⁷⁸ This contractarian theory fails to fulfil the condition of generality.

Moreover, Harsanyi's conception of practical reason competes with the rules that govern our practical reasoning: rationality and reasonableness (here rationality refers to the instrumental conception of rationality, but not Harsanyi's conception of

⁷⁷ The relationship between these fundamental interests and human nature is also mentioned by Rawls. See Rawls (1999a: 160).

⁷⁸ Indeed, Harsanyi himself also acknowledges that this moral preference is rare and hardly influences the behaviours of people. However, it seems that he overlooks the negative effect of this fact to his theory. See Harsanyi (1982: 47).

rationality). Indeed, from the perspective of actual people, the principle of average rule-utilitarianism is neither rational nor reasonable. It is irrational because the choice of hypothetical contractors is made behind the veil of ignorance. The choice of the principle of average rule-utilitarianism makes sense when individuals know nothing about their particular information. However, actual people are fully informed. Their rational deliberation is highly different from the rational deliberation of hypothetical contractors in Harsanyi's contract. Thus, actual people may rationally choose other alternative principles. The principle of average rule-utilitarianism is unreasonable as well. Reasonableness requires people to take the claims of others seriously. People should respect one another as creatures that are capable of asking for justification, 'a non-derivative source of reason'.⁷⁹ This entitlement to respect is absolute and cannot be infringed because of aggregate social welfare. However, this entitlement to respect is exactly the thing which is ignored in the principle of average rule-utilitarianism. According to this principle, the government should merely aim at maximizing average utility; even some people think that this is seriously unjustified. People under this principle are merely taken as a means to maximize average utility, but not ends that can ask for justification. Hence, Harsanyi's hypothetical contract also fails to fulfil the condition of priority, since both rationality and reasonableness justify obligations that go against the principle of average rule-utilitarianism.

Surely, it is going too far to say that Harsanyi's rule-utilitarian theory is a bad moral theory. Harsanyi could respond to my objection by clarifying that he intends to offer a theory of content, but not a theory of obligation and authority. The practical reasoning of his hypothetical contractors does not reveal the practical reasoning of ordinary people. Rather, it merely shows an objectively correct way of deriving moral principles. If actual people deny their obligation to these moral principles, then it is because of their weakness and selfishness, but not the flaw of his contract theory.

⁷⁹ Scanlon (2008: 92).

Although Harsanyi's theory can be understood in this way, his theory is still dissatisfactory when it is evaluated from a contractarian perspective. As I explained before, contractarianism aims at justifying normative principles and explaining why people ought to follow these principles. These aims are also widely accepted by contemporary contractarians. However, Harsanyi's theory merely lays out a set of moral principles which are derived from a hypothetical contract story. It fails to offer a persuasive account of the relationship between the hypothetical contract and the actual people. Thus even though Harsanyi's theory may not be a bad moral theory, it is nevertheless a bad *contractarian* theory.

In conclusion, fulfilling the condition of generality justifies *the imposition of political principles*. The condition of generality ensures that the practical reasoning of the hypothetical contractors is generally understandable to actual people. On the other hand, fulfilling the condition of priority justifies *the overridingness of political principles*. The condition of priority ensures that the hypothetical contractors take all the rules that govern practical reason into account.

2.6 Conclusion

Contractarianism is a methodology which attracts a lot of advocates in contemporary political philosophy, yet it also attracts various critiques. Gauthier's 'Why Contractarianism?' is an attempt to defend contractarianism as an attractive methodology. I appreciate his ambition but I do not think his discussion goes deep enough, since he puts too much attention on only one of the virtues of contractarianism (the virtue of naturalism) and does not take critiques of contractarianism seriously enough. Hence, in this chapter I took up the aims of Gauthier's article but pushed the discussion deeper. I suggested three virtues of contractarianism in order to explain its attractiveness to philosophers. First, contractarianism justifies political principles by the subjective will of individuals

without relying on non-natural, metaphysical properties which conflict with the prevailing scientific worldview. Secondly, contractarian justification shows respect for the will of the individual without taking them merely as means to achieve certain goals. Thirdly, the assumption of contractarianism is more realistic because it does not assume that members of society share a substantial consensus on political affairs. Clearly this list does not exhaust all the virtues of contractarianism, but it goes some way to explaining why contemporary contractarians insist on using this methodology. By adopting this methodology, contractarians can avoid the problems of natural law theorists, utilitarians and communitarians, and justify political principles in a more secular, acceptable and realistic approach.

I have also discussed the standard critique of contractarianism in this chapter. Many critics argue that the consent of hypothetical contractors cannot have normative force for actual people, so it is unclear why actual people are bound by the hypothetical contract. The standard critique is correct that the hypothetical contract should not be confused with actual contract. However, this critique misunderstands the relationship between hypothetical contractors and actual people. The function of the hypothetical contract is to model practical reasoning. Why actual people should follow the hypothetical contract is not because the hypothetical contractors legislate for them, but rather because the political principles specified by the hypothetical contract are the results of correct practical reasoning. The normativity of the hypothetical contract is justified by practical reason. Through the hypothetical contract, we can clarify what political principles would be agreed when a group of people deliberate upon the question of political principles in accordance with the rules that govern practical reason.

Hence, although it is correct that hypothetical consent *itself* does not have any normative force, it has binding power insofar as the hypothetical contractors can model the practical reasoning of actual people correctly. I then further argue that

Chapter 2 Why Contractarianism?

whether the practical reasoning is modelled correctly depends on two conditions: the condition of generality and the condition of priority. Provided that these two conditions are satisfied, the political principles specified by the hypothetical contract have their root in the practical reason. Actual persons are involuntarily subjected to the normative force of these political principles. Although the problem mentioned in the standard critique really exists, it can be avoided as long as the two conditions are

In conclusion, compared to other schools of thought, contractarianism has several virtues, but it also has the flaw of being detached from the real world. This flaw can only be compensated insofar as certain conditions are fulfilled. In the following chapters, I will examine three contractarian models, which are based on different conceptions of practical reason. Then I will examine whether these three models can fulfil the condition of generality and the condition of priority. If they fail to fulfil either of these conditions, then they are vulnerable to the standard critique of contractarianism and fail to develop a connection between the hypothetical contract and actual people.

Chapter 3 Three Models of Contractarianism

3.1 Introduction

Contractarianism is a methodology rather than a theory; thus there are different ways of employing it. In the last chapter I argued that contractarianism is a heuristic methodology which derives substantial political principles from an abstract conception of practical reason. The decision of hypothetical contractors represents the conclusion of correct practical reasoning. This is the general spirit which is shared among different versions of contractarianism. However, despite this, contractarianisms can differ from each other because of different assumptions. The task of this chapter is to introduce different contractarianisms and discuss the assumptions behind them.

In general, a contractarian theory consists of the following three elements:

- (1) A conception of practical reason
- (2) Hypothetical contractors
- (3) A hypothetical agreement

The conception of practical reason is a description of our ordinary practical thinking. Our practical reasoning is not arbitrary but rather governed by certain rules. What these rules are depends upon a contractarian's understanding of the practical reasoning of human beings. By defining the conception of practical reason in terms of these rules, contractarians describe a correct way of reasoning of a person.

Contractarians have to define hypothetical contractors who can model this way of practical reasoning. These contractors have specific characteristics and are placed in specific circumstances so that they think in accordance with the rules that govern practical reason and are not affected by irrelevant factors. Their reasoning represents a 'pure' form of practical reasoning. Contractarians then describe a 'state of nature', that is, a hypothetical situation in which these hypothetical contractors deliberate about how their society should be organised. These contractors discuss with one another in order to reach an agreement. Obviously this state of nature has never before existed in history. It merely helps us clarify what political principles would be accepted in a society if every person could reason correctly. Finally, these hypothetical contractors will agree on certain political principles. This agreement will govern the basic structure of the society, which has a profound influence on the life prospects of the society's members. The distribution of fundamental rights and duties, and the division of the advantages of social cooperation are determined by this agreement. Contractarians then answer the philosophical question of how the society in the real world ought to be organised by adducing this hypothetical agreement.

These three elements constitute the general structure of a contractarian argument. Versions of contractarianism share this structure, and fill in this structure with different elements. Definitions of these three elements are not independent of one another. First, the content of the hypothetical agreement depends on how one defines the hypothetical contractors. Secondly, hypothetical contractors' characteristics and circumstances depend on what conception of practical reason the contractarian assumes. Hence, the conception of practical reason is the most fundamental assumption of the whole contractarian argument. It determines definitions of all the other components.

There are usually two conceptions of practical reason with which contractarians are concerned. One is the conception of rationality; and the other is the conception of reasonableness.¹ Based on these conceptions, two models of contractarianism are possible, each of which may be associated with a different historical figure in the contract tradition. The model which is based on the conception of rationality and presents the hypothetical contract as an agreement among individual utility-maximizers is usually called Hobbesian contractarianism. The Kantian contractarianism model is based on the conception of reasonableness and presents the hypothetical contract as an agreement among reasonable persons who respect one another. Although this classification is widely adopted,² it is incomplete, since it overlooks a distinctive contractarian model which cannot be categorised into either of these two strands. This third model is based on an assumption that people have a *dual* conception of practical reason, that is, people are both rational and reasonable. Since this model has the features of both Hobbesian and Kantian contractarianism, this model can be called *hybrid* contractarianism. In the following section, I will discuss these contractarian models separately and show how these models are developed from their assumptions regarding practical reason.

¹ The definition of rationality should be clarified here. Rationality usually has two definitions, a broad definition and a narrow definition. The broad definition refers to a general name for all capacities for reasoning of human beings, including theoretical reason and practical reason. Practical reason is only a part of rationality. The narrow definition refers to the instrumental conception of rationality, which defines practical reason in terms of several commonly accepted rules, such as the mean-ends rule. This instrumental conception of rationality is only one of the interpretations of practical reason. In this thesis, I adopt the narrow definition of rationality. Hence, rationality and reasonableness represent two conceptions of understanding practical reason.

² See Hampton (1993a); Hampton (1991); Hampton (1993b); Freeman (1990); Freeman (2006); Hamlin (1989); Darwall (2003b); Watson (1998); Buchanan (1990) and Sayre-McCord (2000). Brian Barry's categorization is also relevant, although not explicitly about contract theory. See Barry (1989).

3.2 Hobbesian contractarianism and the conception of rationality

Hobbesian contractarianism takes the conception of rationality to be the conception of practical reason and proposes that a mutually advantageous agreement is rationally justified. This strand of contractarianism includes David Gauthier, James Buchanan, Gregory Kavka, Jean Hampton and Gilbert Harman.³ This is not to deny that different Hobbesian contractarians adopt different assumptions and conclusions. Since it is impossible for me to discuss all of them, I will choose the most plausible version; the Hobbesian contractarian model of David Gauthier. In this section, I will explain how Gauthier defines rationality and constructs a contractarian model on the basis of this conception of practical reason.

3.2.1 The conception of rationality

The conception of rationality is initially attributed to *Homo sapiens* in virtue of their capacity for reasoning and acting upon the result of deliberation.⁴ To say that an individual person is rational is to say that this person measures up to a minimal standard of competence. The origin of the modern study of rationality can be traced back to Hobbes and Hume. According to Hobbes, something can be called good for people only if it is an object of their desire.⁵ A person usually has various desires.

³ See Gauthier (1986), Buchanan (1977), Hampton (1986), Kavka (1986) and Harman (1977). In fact, Kavka's contract theory is not a typical Hobbesian contractarian theory. Rather, Kavka attempts to combine Hobbesian contractarianism and Rawlsian contractarianism together. However, since Kavka emphasizes throughout the book that human beings are generally motivated by their self-interests, and he has a significant contribution to the development of Hobbesian contractarian theories, he should be treated as a Hobbesian contractarian. For the discussion of the 'hybrid' character of Kavka, see Kraus (1993: 204-215).

⁴ Miller (1987: 419-420).

⁵ Hobbes (1994: 15-17).

Reason comes into account when a person has conflicting appetites and aversions.⁶ Hobbes argues that there are ‘general rules’ of action which apply to all persons and which can be found by reason. By following these rules, persons can satisfy their desires effectively. People are creatures pulled by desire, and reason serves as an instrument for the satisfaction of desires. As he famously writes, ‘Reason is the pace; increase of science, the way; and the benefit of mankind, the end’.⁷ Like Hobbes, Hume also holds that actions are produced by two elements: passion and reason, with reason as merely a means which guide us to satisfy our passions. ‘Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them’.⁸

Nevertheless, the descriptions of rationality of Hobbes and Hume are too simplistic. Hence, Hobbesian contractarians borrow a more sophisticated conception of rationality from the contemporary rational choice theory. To rational choice theorists, rationality is a formal procedure which ranks preferences in terms of utility and maximizes individual utility. This conception of rationality is formed by three fundamental conceptions: preference, utility and maximization. First I will explain the conception of preference. When a person is motivated to do something, it can be seen as a preference, which is a consideration related to a state of affairs. A preference refers to an intention that a person has when they think that a particular state of affairs is valuable and they want to achieve this state of affairs: ‘one speaks of preferring an apple to a pear, but more strictly one prefers the eating of an apple to the eating of a pear in some given environment’.⁹ When this state of affairs is realised, then this preference is satisfied. Hobbesian contractarians argue that human

⁶ Hobbes (1994: 17-18).

⁷ Hobbes (1994: 26).

⁸ Hume (2000: 266).

⁹ Gauthier (1986: 22).

beings are animals of preferences and their lives are a journey of satisfying various preferences.¹⁰

However, one has only limited time and resources, so one has to order one's their preferences and choose to satisfy certain preferences first. To this end, some preferences are more important than others. Therefore we need the idea of utility to act as a measure. Utility is an ordinal measure of preference which is 'defined over the possible outcomes in any choice situation if and only if preference weakly orders those outcomes, from most preferred to least preferred'.¹¹ A rational person assigns a utility to each of their preferences. In order to achieve a consistent order of preferences, the assignation of utility has to be governed by several principles of coherence. For example, a person must follow the principle of completeness when they rank their preferences, that is, 'for any two possible outcomes in a choice situation, the chooser must either prefer one to the other or be indifferent between them...it rules out preferentially non-comparable outcomes'.¹² Another example is the principle of transitivity, which requires a person not to rank their preferences as a cycle, that is, they cannot prefer A to B, prefer B to C, then prefer C to A. 'Such cycle would divest choices of all rationality'.¹³ By ranking preferences in accordance with these principles of coherence, a person can get a coherent order of preferences which is measured by utility. The higher the utility is, the more desirable the preference is.¹⁴ As Gauthier notes,

¹⁰ Gauthier (1977: 338-340).

¹¹ Gauthier (1986: 39).

¹² Gauthier (1986: 39).

¹³ Gauthier (1986: 41).

¹⁴ Actually, the individual utility refers to the *expected utility* here. The expected utility does not only depend on how much an individual desires a state of affairs, but also depends on the probability of achieving this state of affairs. By multiplying the utility by its probability, we can get *expected utility*, and it is this, rather than just utility judgments alone, that affects our decision-making. This can gain support from the fact that one outcome being preferred to another is not enough in itself to make us

Chapter 3 Three Models of Contractarianism

Given a set of [relations of preference], constituting the preferences of one individual in a choice situation, then a utility is to be assigned to each possible outcome in such a way that one may infer the person's preference between any two outcomes from the utilities. For any two possible outcomes the one with greater utility must be preferred.¹⁵

It is worth noting that the preferences and utilities of an individual are merely subjective: 'values are ascribed to states of affairs, the ascription is attitudinal, not observational, subjective, not objective'.¹⁶ Whether a state of affairs is valuable is only relative to particular individuals. While one might desire a particular state of affairs, the other might have no interest in it. Even if everyone is rational and given full information, a person's preference might still differ from others'. 'Each person's preferences determine her values quite independently of the values of others'.¹⁷

The idea of preference explains why an individual is motivated to do something, and the idea of utility explains how individuals rank their preferences. After ranking the preferences by utility, rational people should maximize their individual utility; 'Practical rationality in the most general sense is identified by maximization'.¹⁸ Practical reason itself cannot motivate an individual to act. Also, it cannot tell individuals what they should prefer. All these things are out of the control of practical reason.¹⁹ What practical reason is concerned with is only how to maximize utility. To this end, practical reason acts only as an instrument to find a way that satisfies preferences better than any other available alternatives. 'The rational actor maximizes her utility in choosing from a finite set of actions, which take as possible

pursue the former rather than the latter. For successfully securing the latter outcome may be much more probable than securing the former. However, since this point is not so relevant, I will continue to use individual utility to represent the utility function of an individual, instead of using expected utility. For the discussion of expected utility, see Gauthier (1986: 43).

¹⁵ Gauthier (1986: 23).

¹⁶ Gauthier (1986: 25).

¹⁷ Gauthier (1986: 25).

¹⁸ Gauthier (1986: 22).

¹⁹ Gauthier (1986: 26).

outcomes the members of a finite set of states of affairs'.²⁰ She will concentrate on making good use of her finite time and resources, seeking the most effective way in order to achieve the highest possible utility. Gauthier articulates this disposition as such: 'The rational man is...simply the man who seeks *more*'.²¹

In general, rationality can be seen as a 'weighing model'. It solves the conflicts among different preferences by weighting their importance to the person. People can rank preferences by individual utility and satisfy them in a way which maximizes their individual utility. Moreover, the aim of this formal procedure is to look for a way of living well. It aims at the satisfaction of preferences and believes that, by satisfying more preferences, a person's life can improve. Putting it in another way, it takes the notion of good as the fundamental notion and defines the notion of right as maximizing individual good.

3.2.2 Hobbesian contractarianism

This conception of practical reason is a suitable ground for developing contractarian theory because it is a process of reasoning that is widely accepted.²² The conception of rationality does not presuppose any substantial goals.²³ It is impartial with respect to all goals and simply posits people pursuing these goals consistently. The hypothetical contract of Hobbesian contractarianism is to illustrate what agreement would be made if every citizen could reason in accordance with this conception. I will leave the detailed discussion of the hypothetical contract to Chapter 4, which

²⁰ Gauthier (1986: 22).

²¹ Gauthier (1977: 344).

²² The general acceptability of the conception of rationality will be further explained in Section 4.3. Also, I do not intend to claim that all kinds of practical reasoning can be explained by the conception of rationality, since rationality is merely one of the aspects of practical reason. For example, rationality cannot explain the kinds of reasoning that the idea of obligation plays a crucial role.

²³ Broome (1991: 90-92).

discusses Hobbesian contractarianism in-depth, for this section aims only to explain how rationality can be used to ground contractarian theory. This section will discuss the characteristics of hypothetical contractors and the hypothetical contract of Hobbesian contractarianism only briefly.

After adopting rationality as the conception of practical reason, the next task of Hobbesian contractarians is to clarify what political principles would be agreed upon if all people were rational. Hobbesian contractarians propose a 'state of nature' in which a group of well-informed utility-maximizers bargain with each other. These utility-maximizers have to agree on certain political principles in order to exit the non-cooperative state in which everyone is worse-off.²⁴ One of the famous descriptions of the state of nature is that of Hobbes, who argues that, without a system of law enforcement, each person fears others as potential attackers and may also attack others in order to remove potential future threats. Therefore, a state of nature is a war of all against all. Hobbesian contractarians usually do not adopt this negative picture, but they all agree that the non-cooperative state is undesirable. Hence the Hobbesian contract story is of how a group of well-informed utility-maximizers cooperate with each other in order to leave this undesirable state.

Since all hypothetical contractors are utility-maximizers, they care most about how to maximize their utility in the agreement. However, everyone also understands that others would not agree if the agreement only favoured the pursuit of his utility. Hence, not surprisingly, the agreement will be political principles that are mutually advantageous. These principles should guarantee a Pareto-optimal outcome, an outcome where everyone's utility is maximized in a way compatible with each other.²⁵ Such an agreement brings all hypothetical contractors out of this mutually

²⁴ Kavka (1986: 97-101) and Hampton (1986: 58-79).

²⁵ Gauthier (1986: 76).

disadvantageous situation and coordinates them to be their mutually advantage, that is, compared with the state of nature, all gain a higher utility. Hobbesian contractarians then further argue that this hypothetical agreement is justified to actual people as well. Since the practical reasoning of actual people is also governed by rationality, they would agree with political principles if they recognised that this agreement is rationally justified.²⁶

We can now conclude our exposition of Hobbesian contractarianism. The three elements of Hobbesian contractarianism are: (1) the conception of rationality, (2) hypothetical contractors that are well-informed utility-maximizers and, (3) a hypothetical agreement that is mutually advantageous. Among these three elements, the second and third elements depend on the first. The conception of rationality determines the remaining parts of the whole Hobbesian contractarian argument. Hobbesian contractarians make this assumption because rationality is a weak assumption that people widely accept. The generality and formality of rationality may be able to explain why Hobbesian contractarianism is so intuitively appealing to many people: given that we always prefer having higher utility, why resist a principle which could bring us higher utility?

3.3 Kantian contractarianism and the conception of reasonableness

Rationality is a widely accepted interpretation of practical reason, but it is not the only interpretation. Some contractarians make a different conception of practical reason. This assumption is the conception of reasonableness. The second strand of contractarianism is in many ways the opposite of the first. Since contractarians of this strand take Kant as their predecessor, this strand is usually called Kantian

²⁶ See, for example, Buchanan (1977: 54), Gauthier (1986: 11) and Hampton (1986: 56).

contractarianism. In contrast with Hobbesian contractarianism, Kantian contractarianism interprets practical reason in terms of reasonableness, and proposes a fair agreement which could be justified to everyone.²⁷ This strand of contractarianism includes T. M. Scanlon, Brian Barry, Thomas Nagel, David Richards and Stephen Darwall.²⁸ In this section I will introduce Kantian contractarianism and focus particularly on the most impressive and influential version of it that has been advanced by T. M. Scanlon.

3.3.1 The conception of reasonableness

Both rationality and reasonableness derive from the same Latin root *ratio*.²⁹ However, unlike rationality, reasonableness presents a different picture of practical reason. It is a ‘testing’ model, a procedure which takes certain constraints as absolute and tests desires by this constraint. Hence it does not focus upon which desires are more important to a person or how to rank these desires in order to lead a good life. Rather, it focuses only on whether or not these desires violate the absolute constraint. The

²⁷ Some commentators argue that this strand of contractarianism presupposes certain pre-existing moral rights and duties. The purpose of the social contract is to protect these pre-existing moral rights and duties. See Boucher and Kelly (1994b: 4). However, this is a misunderstanding, because Kantian contractarians take rights as the ‘product’ of contract rather than take them as the presupposition of contract. As Scanlon argues, what rights people have depends on reasonableness, that is, justifiability to other individuals. See Scanlon (2003a: 3).

²⁸ See Scanlon (1998), Barry (1995a), Nagel (1991), Richards (1971) and Darwall (2009). Some might wonder why Richards is included in this family, since Richards does not mention the conception of reasonableness and defines moral principles as principles that would be rationally chosen in a fair situation. However, when Richards explains why actual people have to take his hypothetical contract seriously, he argues that ‘as a brute fact of human psychology, there is a widespread desire to be moral.’ (Richards 1971: 242) According to his discussion, this desire is in fact a desire to treat others fairly, that is, a desire to be reasonable. Hence, although the conception of reasonableness does not appear in his writings, it actually plays a crucial role in his theory.

²⁹ Miller (1987: 420).

constraint in the conception of reasonableness is the ‘constraint of justifiability’.³⁰ It is a constraint that requires us to take public justifiability as a criterion for evaluating our behaviour. Kantian contractarians presuppose that reasonable persons share an aim of living with other reasonable persons.³¹ Every reasonable person desires to cooperate with others under fair terms of social cooperation.

In the contractualist analysis of right and wrong, what is presupposed first and foremost is the aim of finding principles that others who share this aim could not reasonably reject. This aim then brings other reasons in its train. Given this aim, for example, it would be unreasonable to give the interests of others no weight in deciding which principles to accept. For why should they accept principles arrived at this way?³²

Since reasonable people want to live with other people who are also reasonable, they have to respect other people’s capacity for reason-giving.³³ The way to respect others’ reason-assessing capacity is to take the possible claims of others into account. A reasonable person has to be fair-minded, judicious and able to see other points of view. One should try to put oneself in the shoes of others and should think about whether or not ‘I’ would accept this behaviour if ‘I’ were ‘the other’. Hence one should ‘test’ one’s actions by the constraint of justifiability.

This constraint is an absolute criterion because any desires which violate this criterion should be rejected, no matter how important the desires are. As Scanlon argues, this constraint should not be seen as a consideration which is weighted against other desires. Rather, it is a criterion that *determines* the weight of a person’s

³⁰ This feature of the Kantian conception of practical reason is also suggested by Samuel Freeman, who also argues that, in Kantian conception of practical reason, motivations are tested to see whether they have certain *structure*. See Freeman (1991: 298).

³¹ Scanlon (1998: 154).

³² Scanlon (1998: 192).

³³ Scanlon (1998: 194).

desires.³⁴ If a reason violates this constraint, then it has no weight. That is why reasonableness is a ‘testing model’, which is different from the weighting model represented by rationality. It does not simply take all desires into account and weigh the importance of them. Rather, it sets a constraint as absolute and only those desires which are compatible with this constraint can be taken into account. Whether a desire is compatible with this constraint is more important than how much weight this desire has in a person’s utility function. Some Kantian contractarians believe that this testing model is a more phenomenologically accurate account of normative reasoning.

It is, phenomenologically, much more plausible to suppose that, certainly for the fully moral person and even for most of us much of the time, these considerations are excluded from consideration well before the stage at which we decide what to do. Being moral involves seeing reason to exclude some considerations from the realm of relevant reasons (under certain considerations) just as it involves reason for including others. The contractualist account can explain this fact, since these considerations are ones that others could reasonably refuse to license us to count as reasons.³⁵

In light of this absolute constraint, it is unsurprising that reasonableness, in contrast with rationality, does not presuppose that human beings are merely animals with preferences. Rather, it presupposes that human beings will not be governed by preferences unless these preferences ‘can be publicly justified to others according to the system of norms generally accepted within the group’.³⁶ Human beings can control their desires in order to comply with the constraint of justifiability, even at the cost of lowering their individual utility.³⁷ Practical reason is not a slave to

³⁴ Scanlon (1998: 156-157).

³⁵ Scanlon (1998: 157).

³⁶ Freeman (1990: 22). For the examples of the moral-based contractarian who argues that practical reason can be independent of rational maximization of utility, see Scanlon (1988: 173).

³⁷ See Darwall (2006) and Freeman (1990: 24).

passions, for it can reject passions if they are unreasonable. ‘We judge and act autonomously...by proceeding from and regulating our activities according to certain principles’.³⁸ It shows that the aim of practical reason is to seek a way of living right. This conception of practical reason takes the notion of right as fundamental and restricts the pursuit of individual good within the limit of rightness.

3.3.2 Kantian contractarianism

Like rationality, reasonableness is also an important aspect of practical reason. People usually have a sense of respect for one another and this can be seen by the sense of guilt that appears when people find that they are treating others in an unjustifiable way. The generality of reasonableness explains why Kantian contractarians take it as the conception of practical reason. This assumption determines the characteristics of hypothetical contractors and the content of hypothetical agreement in Kantian contractarianism. Again I will leave the detailed discussion of these two elements to Chapter 5, which discusses Kantian contractarianism specifically, because this section is merely to illustrate the relationship between Kantian contractarianism and the conception of reasonableness. Let me briefly explain here these two elements in Kantian contractarianism.

In order to derive political principles from the conception of reasonableness, hypothetical contractors are defined as well-informed reasonable persons who mutually respect one another. Kantian contractarians ask us to imagine a situation in which a group of reasonable persons deliberate together for the sake of achieving an agreement on political principles.³⁹ Since they are reasonable, they will not propose

³⁸ Freeman (1990: 42).

³⁹ Scanlon (2003a: 5).

principles which could not be justifiable to any of them. For example, they would not propose principles that favour certain people arbitrarily. This is because, if they were people who were disadvantaged arbitrarily, then they would not agree with these principles.⁴⁰ Hence contractors will only propose principles which could be publicly justifiable to all.

However, there is more than one principle which could be justifiable to everyone. Publicly justified principles can be incompatible with one another. So reasonable persons have to give reasons to justify their proposals and explain why the proposals of others should be rejected. The process of exchanging reasons will go on. Numerous proposals will continue to be suggested and rejected. Consequently, reasonable persons will get political principles which are supported by the strongest, the most persuasive reasons.⁴¹ They are 'principles which no one could reasonably reject'.⁴² Kantian contractarianism then proposes that a justified state should be governed by these principles. Even though real people have not actually agreed with these principles, they would agree with them since their practical reasoning is, like that of hypothetical contractors, also guided by reasonableness. 'A legitimate government is one whose authority citizens can recognise while still regarding themselves as equal, autonomous...agents'.⁴³

⁴⁰ Scanlon (1998: 216). See also Scanlon (1977: 69).

⁴¹ Some people, such as Jean Hampton, might object that this is not a 'contract' because no bargaining is involved. There is no bargain among people who have different perspectives, since all people are holding a common perspective. Even a single deliberator can also arrive at the same principles. See Hampton (1993). Although Hampton is criticizing Rawls in this article, this critique can be applied to Kantian contractarianism as well. However, it is too narrow to say that an undertaking of contract must involve bargaining among people. In fact, a contract can also be an agreement that a group of people jointly commit to something which is publicly justified to them. The agreement of Kantian contractarianism is certainly this kind. See Freeman (1990: 35).

⁴² Scanlon (1998: 213-218).

⁴³ Scanlon (1972: 14).

We can now conclude our exposition of Kantian contractarianism. The three elements of Kantian contractarianism are: (1) the conception of reasonableness, (2) hypothetical contractors that are well-informed reasonable persons who mutually respect one another, and; (3) a hypothetical agreement that no one could reasonably reject. Like Hobbesian contractarianism, the second and third elements depend upon the first; the conception of reasonableness. Kantian contractarians believe that, compared with the conception of rationality, the conception of reasonableness offers a better account of practical reason and thus, offers a better ground for a contractarian theory.

3.4 The myth of dichotomy

In the last two sections we saw the two most commonly seen contractarian models. Both of them are based on widely accepted conceptions of practical reason. Hobbesian contractarianism is based on the conception of rationality, while Kantian contractarianism is based on the conception of reasonableness. Rationality and reasonableness emphasise different aspects of human nature. Rationality emphasises the passive side of human nature, our capacity to be pleased and satisfied. In this description, we are attracted to various preferences, and are finally motivated by the most attractive. Reasonableness emphasises the active side of our human nature; our capacity for control and constraint. In this description, we can stand aside from these preferences, evaluate them by the constraint of justifiability and resist temptation by them if they violate this constraint. Here different ethical concerns can be identified. The former is more concerned about the notion of good, while the latter is more concerned about the notion of right. Also, from these different perspectives on human nature, they deliberate upon practical issues in different ways. The former uses a model that weights preferences by individual utility and aims at maximizing individual utility. The latter uses a model that tests preferences by the constraint of

justifiability.⁴⁴ The differences between these two conceptions of practical reason are set out in the table below:

Table 1 Rationality and Reasonableness

	Rationality	Reasonableness
Agency	Passive	Active
Aim	In the pursuit of good	Acting according to right
Model of deliberation	Weighing model, ranks preferences in a way which can maximize individual utility	Testing model, makes sure that preferences do not violate the constraint of justifiability
Contractarianism	Hobbesian contractarianism	Kantian contractarianism

⁴⁴ Surely it does not mean that the conception of rationality does not involve any ideas of constraint. For example, the budget constraint, which has always been mentioned in the standard consumption theory, plays a crucial role. I am indebted to Albert Weale for this point.

It seems that the dichotomy between Hobbesian and Kantian contractarianism is exclusive and other possible contractarianisms will only be similar to Harsanyi's contractarianism, which is based on a supererogatory conception of practical reason. However, I will argue that this dichotomy is too simple because it fails to include those contractarians who belong to neither of them. Some contractarians, such as Rawls, develop a distinctive model which has characteristics of both of these contractarianisms.⁴⁵ In this section, I will explain why Rawls' contractarian theory does not belong to either side and how the dichotomy of contractarianism does not reveal fully the complexity of certain contractarian theories.

3.4.1 A re-interpretation of Rawls' contractarianism

Before I discuss Rawls' contractarian theory, there is a misunderstanding that should be noted. People usually take the original position, a situation in which hypothetical contractors choose behind a veil of ignorance, to be the hypothetical contract of Rawls. This is not a sound interpretation of Rawls' contractarian theory because this interpretation is vulnerable to many critical objections. For example, because of the veil of ignorance, there is a great gap between the reasoning of hypothetical contractors and actual people. It is questionable why actual people should take the decisions of hypothetical contractors seriously, for these decisions are made with very limited information.⁴⁶ These critiques can be avoided if we adopt another interpretation of Rawls' contract theory. We should instead understand Rawls' social contract as describing a public agreement in a well-ordered society.⁴⁷ In this interpretation, the hypothetical contractors are not mutually disinterested parties

⁴⁵ Although Rawls emphasises that his principles of justice are principles that are applied to the basic structure, he does not deny that individuals also have obligations to these principles. The obligations of individual 'are an essential part of any theory of justice'. Hence, like Gauthier and Scanlon, Rawls also has to explain why actual people are bound by his hypothetical contract. See Rawls (1999a: 237).

⁴⁶ Brown (1988: 444).

⁴⁷ Freeman (2007a: 4-5).

behind the veil of ignorance,, but rather well-informed rational and reasonable people. The hypothetical agreement is not the agreement made in the original position, but rather the agreement publicly acknowledged in the well-ordered society. I will discuss further about the advantage of this interpretation in Chapter 6, which discusses Rawls' contractarianism specifically. But at this stage we need only pay attention to the characteristics of hypothetical contractors in Rawls' contract.

The hypothetical contractors in Rawls' theory are free and equal persons who have two moral powers: 'first they are capable of having (and are assumed to have) a conception of their good (as expressed by a rational plan of life); and second they are capable of having (and are assumed to acquire) a sense of justice, a normally effective desire to apply and to act upon the principles of justice'.⁴⁸ Since these two capacities refer to rationality and reasonableness in our practical reason respectively, Rawls' contractarian theory has the features of both Hobbesian and Kantian contractarianism.

3.4.2 The conception of reasonableness in Rawls' contractarianism

I will begin by examining the similarity between the capacity for a sense of justice and reasonableness. First, although Rawls is arguably rather vague in defining this motivation, the sense of justice is a testing model which rejects interests that are incompatible with certain absolute evaluative criteria. Being reasonable means looking for a way of living in accordance with the constraint of justifiability. The constraint of justifiability acts as a principle of 'right', which regulates the 'good' of individuals, since it has an absolute priority over the pursuits of all goods and values. This requirement of priority can also be found in the sense of justice. According to

⁴⁸ Rawls (1999a: 442).

Rawls, the sense of justice is ‘a regulative desire to act upon certain principles of right’.⁴⁹ People who have an effective sense of justice accept ‘the main idea...that given the priority of right, the choice of our conception of the good is framed within definite limits. The principles of justice and their realisation in social forms define the bounds within which our deliberations take place’.⁵⁰ To them, the conceptions of the good which, ‘require the violation of justice have no value. Having no merit in the first place, they cannot override [justice’s] claim’.⁵¹ Hence, they are willing to honour these principles ‘even at the expense of their own interests as circumstances may require, provided others likewise may be expected to honour them’.⁵² Therefore, although Rawls does not use the phrase ‘testing model’, the sense of justice effectively describes the testing model of the conception of reasonableness.

Secondly, the sense of justice also requires people to regulate themselves in a way which could be justifiable to others. The sense of justice is always misunderstood as merely representing an empty sense of duty towards moral principles. But, according to Rawls, the sense of justice is not merely an empty motivation. It ‘expresses a willingness...to act in relation to others on terms that they also can publicly endorse’.⁵³ This willingness implies an acknowledgement of the ‘burdens of judgment’. Free and equal persons must acknowledge that others inevitably affirm conceptions of the good different from their own.⁵⁴ If they want to cooperate with others, then they must respect others and cooperate in the terms that could be publicly justifiable to others.⁵⁵ That is why they have to honour principles of justice, since these principles represent fair terms of social cooperation that take the claims

⁴⁹ Rawls (1999a: 491).

⁵⁰ Rawls (1999a: 493).

⁵¹ Rawls (1999a: 28).

⁵² Rawls (2001: 7).

⁵³ Rawls (1993: 19).

⁵⁴ Rawls (1993: 54-58).

⁵⁵ Rawls (1993: 81-82).

of each person into account.⁵⁶ Honouring these principles represents living in a way that could be justifiable to each person.⁵⁷ From this we see that the constraint of justifiability, which is the constraint presupposed in the conception of reasonableness, plays a crucial role in the sense of justice.

From these two stances, one can see the similarities between Rawls' and Kantian contractarianism. Both of them are based on a conception of practical reason that takes the constraint of justifiability as an absolute evaluative criterion. Fairness and respect, which are the ideas implicit in the constraint of justifiability, also, 'shape Rawls' thought at the deepest level'.⁵⁸ That is why the contents of the hypothetical agreements of these two contract theories resemble one another. Since the hypothetical contractors in Rawls' theory are reasonable, they would not propose principles which could not be justified to others. Hence they should finally agree upon principles of justice as fairness, which could be publicly acknowledged and could give justification to each citizen's reason. These principles are, as Scanlon acknowledges, effectively positioned whereby no one could reasonably reject.⁵⁹

3.4.3 The conception of rationality in Rawls' contractarianism

In fact, no one should be surprised to see the resemblance between Kantian contractarianism and Rawls' theory. This resemblance is widely recognized by both Rawls and Kantian contractarians.⁶⁰ However, the hypothetical contractors in Rawls' contractarianism also have another characteristic, which is often overlooked by commentators. Apart from reasonableness, they have a capacity for the conception of

⁵⁶ Rawls (1993: 89).

⁵⁷ Rawls (1993: 52).

⁵⁸ Larmore (2003: 391).

⁵⁹ Scanlon (1998: 243-244).

⁶⁰ Rawls (1993: 49n. 2) and Scanlon (1998: 243-245).

the good, which entails that they want to realise their conception of the good rationally. This rational capacity is related to the idea of good. As Rawls said in the section ‘Goodness as Rationality’, the goodness of a life is defined by whether or not it is rational. Rawls first defines rationality in a formal way, like the way adopted by Hobbesian contractarians. Rationality is specified by principles of rational choice such as taking effective means to one’s ends, ranking one’s ends in order of priority.⁶¹ Rawls then defines the notion of good as a life plan chosen in accordance with these principles. In light of these principles, people are assumed to be fully informed, to be able to reflect upon their plan critically, and to appreciate its consequences.⁶² This is the ‘thin theory of good’, based on a formal definition of a person’s idea of the good.

However, this definition of good is too formal, for it says nothing about the ends of a life plan. Rawls believes that a theory of *good* cannot avoid discussing this issue.⁶³ Therefore, he develops a ‘thick theory of the good’. A good life plan not only has to be compatible with the formal principles of rational choice, but also has to be relevant to the highest order of interests of human beings, that is, the interests in realising the two moral powers.

We take moral persons to be characterized by two moral powers and by two corresponding highest-order interests in realizing and exercising these powers. The first power is the capacity for an effective sense of justice, that is, the capacity to understand, to apply and to act from (and not merely in accordance with) the principles of justice. The second moral power is the capacity to form, to revise, and rationally to pursue a conception of the good. Corresponding to

⁶¹ Rawls (1999a: 358-365).

⁶² Rawls (1999a: 365-372).

⁶³ Rawls (1999a: 372).

Chapter 3 Three Models of Contractarianism

the moral powers, moral persons are said to be moved by two highest-order interests to realize and exercise these powers.⁶⁴

These two moral powers constitute the free and equal nature of human beings. People are free because they have the second moral power to reflect upon their conception of the good,⁶⁵ and they are equal with one another because everyone has these two moral powers.⁶⁶ Therefore, realising these two powers amounts to realising the nature of the human being as free and equal. Rawls believes that the interests in realising these two powers are the ‘highest-order’ because ‘these interests are supremely regulative as well as effective. This implies that, whenever circumstances are relevant to their fulfilment, these interests govern deliberation and conduct’.⁶⁷ The highest-order interests should ‘be the ends and activities that have a major place in rational plans’.⁶⁸

This can explain why Rawls’ hypothetical contractors will choose principles of justice as fairness, for these can effectively protect highest-order interests.⁶⁹ Principles of justice as fairness assign several basic liberties an absolute weight, notably the liberty of conscience and equal political liberties.⁷⁰ These basic liberties are necessary conditions for ‘the adequate development and full exercise of the two moral powers of citizens as free and equal persons’.⁷¹ Liberty of conscience enables people to revise their present way of life and change to endorse other conceptions of the good. The equal political liberties enable people to develop and to exercise the capacity to evaluate the basic structure of society. The absolute priority of these basic

⁶⁴ Rawls (1980: 312).

⁶⁵ Rawls (2001: 23-24).

⁶⁶ Rawls (1999a: 442-443).

⁶⁷ Rawls (1980: 312). Similar expression can also be seen in Rawls (1999a: 491).

⁶⁸ Rawls (1999a: 379).

⁶⁹ Rawls (1999a: 221).

⁷⁰ Rawls (1993: 294-295).

⁷¹ Rawls (1993: 297). See also Rawls (2001: 45).

liberties can guarantee that the highest-order interests of each person are secured unconditionally. Thus principles of justice as fairness will be chosen, but not other principles, such as the principle of average utility, since other principles would permit cases where the highest-order interests of people would be sacrificed for the sake of a greater net sum of utility.

Hence we can see that, to Rawls, a free and equal person has two reasons to follow principles of justice: one relates to reasonableness, the other relates to rationality. For the first reason, a free and equal person follows principles of justice for they are principles which could be publicly justified. However, following principles of justice does not mean always sacrificing interests. This leads to the second reason to comply with principles of justice. A free and equal person also follows principles of justice because acting from these principles can advance their highest-order interests. By being just, they can realise these two moral powers and enjoy goods which are fundamental to their lives. Therefore it is both rational and reasonable to comply with principles of justice.

3.4.4 Between Hobbesian and Kantian contractarianism

From the assumption of highest-order interests one can see the difference between Rawls and Kantian contractarians. Kantian contractarianism is based on the conception of reasonableness. Hence the hypothetical contractors are defined as merely reasonable. Even if the hypothetical agreement will seriously harm the interests of some hypothetical contractors, these hypothetical contractors will still agree with it provided that this agreement is reasonably justified.⁷² Reasonableness is the only consideration in the agreement-making procedure of Kantian

⁷² Scanlon (1998: 160-163).

contractarianism. However, Rawls obviously has another view of the hypothetical contract. To Rawls, if a hypothetical agreement requires citizens to always sacrifice their conceptions of the good, then it would be unstable, and an unstable agreement would be ‘seriously defective’.⁷³ A satisfactory political principle should not only be reasonably acceptable to its followers, but also consistent with their view of the rational good. In contrast with Kantian contractarianism, Rawls exerts commendable effort showing that his political principles do not harm the interests of followers, but rather lead them to rationally advance their highest-order interests. The rational capacity of free and equal persons and its relation to the idea of good are often overlooked in discussions of Rawls’ theory. This explains why Rawls is often misunderstood as a Kantian contractarian. However, a short anecdote about Rawls might verify my observation.

[In a conversation with Samuel Freeman,] Rawls always referred to justice as fairness as a ‘contractarian’ position. He was opposed to others’ use of Scanlon’s term ‘contractualism’ as a generic term used to refer to justice as fairness...He said contractualism was Scanlon’s own position, and that it was original, distinctive, and in important respects quite different from what he was trying to do.⁷⁴

Freeman states that part of what Rawls had in mind was that Scanlon’s contract theory does not focus enough on whether political principles are compatible with the good of people, while this is a central concern of Rawls’ contract theory.

On the other hand, the assumption of the highest-order interests also shortens the distance between Rawls and Hobbesian contractarians. Clearly Rawls would deny

⁷³ Rawls (1999a: 518). It is important to note that Rawls does not say that an unstable agreement means that it is unjustified. An agreement can be justified, but still unstable.

⁷⁴ Freeman (1991: 36). A similar story can also be found in Nussbaum (2006: 418n. 9), that Rawls describes himself as a social contract theorist but not a Kantian.

that his theory is indebted to Hobbes, since he claims that his theory ‘generalizes and carries to a higher level of abstraction the familiar theory of the social contract as found, say, in Locke, Rousseau, and Kant’.⁷⁵ Nevertheless, the resemblance between Rawls and Hobbesian contractarianism should not be overlooked. As Robert Paul Wolff observes, Rawls is clearly affected by the neo-classical economic assumption of utility-maximization.⁷⁶ Both Rawls and Hobbesian contractarians assume that hypothetical contractors want to satisfy some preferences. These preferences can be satisfied by following certain political principles, and this is a crucial reason why contractors should follow these political principles. The only difference between them is that they define rational preferences in different ways. Hobbesian contractarians take any considered and coherent preferences as rational preferences, whereas Rawls further insists that rational preferences have to be compatible with the highest-order interests. Despite this difference, both Rawls and Hobbesian contractarians maintain that following political principles is a ‘good’ to contractors.⁷⁷

Following this analysis, one can see that Rawls’ contract theory not only has some characteristics of Kantian contractarianism, but also has some characteristics of Hobbesian contractarianism. By understanding the complex character of Rawls’ contract theory, the inadequacy of the orthodox dichotomy between Hobbesian and Kantian contractarianism becomes evident. In the dichotomy, contractarians are classified into two kinds; either Hobbesian or Kantian. However, Rawls falls into neither camp. More accurately, he is in both. The hypothetical contractors in Rawls’ theory are both rational and reasonable. Hence Rawls is not a Hobbesian

⁷⁵ Rawls (1999a: 10).

⁷⁶ Wolff (1977: 208-209).

⁷⁷ In fact, the Hobbesian character is much stronger in Rawls’ earlier writings. In ‘Justice as Fairness’ (Rawls 1958), Rawls assumes that people are rational, mutually self-interested, and ‘sufficiently equal in power and ability to guarantee that in normal circumstances none is able to dominate the others’, then he discusses what principles these people will agree on. This is highly similar to the Hobbesian contractarian view of the person in the state of nature. Gauthier also discusses several similarities between his theory and Rawls’. See Gauthier (1974).

contractarian, for he assumes that hypothetical contractors are concerned with being reasonable to others. Yet he is not a Kantian contractarian, for he assumes that contractors are also concerned with the rational pursuits of their interests. His contract theory is a 'hybrid' social contract; a hybrid of Hobbesian and Kantian contractarianism.⁷⁸ The orthodox dichotomy fails to include this distinctive kind of contractarian model.

3.5 The third possibility: hybrid contractarianism and the dual conception of practical reason

Since the orthodox dichotomy can only capture 'pure' contractarianism, its inadequacy is that it overlooks hybrid contractarianism, which has characteristics of both Hobbesian and Kantian contractarianism. Furthermore, this kind of model is rare in the history of contractarianism. Rousseau and Rawls belong to this tradition since both of them emphasise that the social contract not only defines principles of

⁷⁸ Here one might recall Brian Barry's analysis of Rawls' theory of justice. Barry also takes justice as fairness as a combination of two theories, justice as mutual advantage, which takes justice as the rational cooperation for mutual advantage under circumstances of justice, and justice as impartiality, which takes justice as a reasonable agreement that is acceptable from all points of view. See Barry (1989: 179-254). Although both Barry and I take Rawls' contractarianism to be a hybrid theory, my analysis is clearly different from Barry's. From Barry's perspective, Rawls is a 'hybrid' theorist merely because he uses the languages of both justice as mutual advantage and justice as impartiality. However, I concentrate on the conception of practical reason behind Rawls' theory and indicate that it is this dual assumption which makes Rawls a 'hybrid' theorist. Moreover, the component of rational advantage understood by Barry in Rawls' theory is the rational bargaining that happened in the original position among parties behind the veil of ignorance. This ignores Rawls' discussion on the 'good' of justice in the Part III of *A Theory of Justice*, which is crucial for understanding why justice as fairness is 'rational' to actual people. The rational interests in the original position should not be confused with the real motivations that motivate free and equal persons to follow principles of justice in the real world. (See Rawls 1980: 320-322, 357-358) In contrast to Barry, I understand rational advantage to be the highest-order interests in realizing two moral capacities. These are also fundamental goods valued by free and equal persons and also the goods that Rawls emphasizes in the Part III of *A Theory of Justice*. When Rawls talks about the congruence between rationality and reasonableness, he is talking about the congruence between highest-order interests and the sense of justice. Hence, I believe that my analysis based on a more correct understanding of Rawls' theory.

‘right’, but also represent the fundamental good of people. As my thesis focuses upon contemporary contractarianism, I will take Rawls as the example of this contractarian model. This contractarianism has different characteristics from that of the other two contractarianisms. Hence it has strengths that these two contractarianisms do not have, while it also has to encounter problems that these two contractarianisms do not need to face.

3.5.1 The dual conception of practical reason

The distinctiveness of hybrid contractarianism originates in its attitude to the relationship between rationality and reasonableness. Different contractarians perceive this relationship in various ways. Hobbesian contractarians subordinate reasonableness to rationality. In this approach, reasonableness becomes one of the many preferences of a person, and it is one of the considerations in the weighting model of rationality. Kantian contractarians subordinate rationality to reasonableness. In this approach, the considerations in the weighting model of rationality must not violate the constraint of justifiability.

Hybrid contractarians take a different approach. In the previous section, we have seen that the hypothetical contractors of hybrid contractarianism have *dual* characteristics: they are both rational and reasonable. This implies that the conception of practical reason of hybrid contractarianism is a *dual* assumption: it assumes that rationality and reasonableness are two aspects of practical reason. Neither is subordinate to the other. The reasonable perspective uses the testing model to assess the justifiability of people’s actions. The rational perspective uses the weighting model to rank the preferences of people by using utility as a measure to maximize utility.

But is this understanding of practical reason plausible? Actually this dual assumption is not uncommon since many theorists have already pointed out that it is inadequate to see human beings as merely rational or reasonable. Rather, rationality and reasonableness should be seen as two aspects of practical reason which are independent of, and irreducible to, the other. For example, David O. Brink argues that the Kantian interpretation of practical reason is inadequate and he proposes a 'dualism of practical reason', which consists of the categorical imperative and the categorical prudence.⁷⁹ The former norm is similar to the conception of reasonableness, whereas the latter norm is close to the conception of rationality.⁸⁰ Brink believes that these two norms have the same categorical nature, and that they thus have the same status in practical reasoning. Both requirements are formal requirements of people's practical reason, so neither of them has supremacy. The requirement of categorical imperative cannot override the demand of categorical prudence, and *vice versa*. Apart from Brink, Elster also suggests that there are two norms in people's practical reason: rationality and impartiality. The norm of rationality, which requires people to arrange their desires according to certain logical rules, corresponds to the conception of rationality discussed earlier, while the norm of impartiality, which requires people to take certain moral rules as absolute, corresponds to the conception of reasonableness. Elster argues that these two norms are independent of each other: 'The many attempts to deduce [rationality] from [impartiality] are doomed to failure'.⁸¹ A good life is a life in which rationality and impartiality are harmonious with each other.⁸² Hence, to Brink and Elster, this dualistic interpretation of practical reason is a more comprehensive and accurate understanding than the tradition Kantian or Hobbesian interpretation. These two interpretations can capture single sides of the picture separately; but not the whole.

⁷⁹ Brink (1997: 289-290).

⁸⁰ Brink (1997: 267-268, 287-288).

⁸¹ Elster (2008: 61).

⁸² Elster (2008: 68).

Although Brink and Elster are not concerned here with contractarianism, it seems that their proposed dualistic interpretation of practical reason is exactly like the conception of practical reason which underlies hybrid contractarianism. Hybrid contractarians, such as Rawls, assume that the practical reason of actual people consists of conceptions of rationality and reasonableness, both having equal status in practical reason. They have different functions and are concerned with different things: one is concerned with the pursuit of good, while the other is concerned with living according to the right. As Rawls emphasises, both capacities are distinctive and neither can be derived from the other:

The reasonable and the rational are taken as two distinct and independent basic ideas. They are distinct in that there is no thought of deriving one from the other; in particular, there is no thought of deriving the reasonable from the rational.⁸³

The concept of justice and goodness are linked with distinct principles. More precisely, each concept with its associated principles defines a point of view from which institutions, actions, and plans of life can be assessed.⁸⁴

The distinction between rationality and reasonableness often appeared in Rawls' writings.⁸⁵ As Samuel Freeman observes, 'there are two ideal perspectives in Rawls' conception of justice: the original position and the deliberative rationality. The former provides the foundation for judgments of justice; the latter provides the basis

⁸³ Rawls (1993: 51).

⁸⁴ Rawls (1999a: 496-497).

⁸⁵ Dreben (2003: 321).

for judgments regarding a person's good'.⁸⁶ This further proves that a dual conception of practical reason lies behind his contractarian theory.⁸⁷

However, one should be careful to note that Rawls, as a hybrid contractarian who is closer to the Kantian tradition, does not think that the relationship between rationality and reasonableness is completely coordinated. In some texts, Rawls emphasises that, 'the Reasonableness presupposed and subordinates the Rationality. It defines the fair terms of cooperation acceptable to all within some group of separately identifiable persons, each of whom possesses and can exercise the two moral powers'.⁸⁸ Although Rawls acknowledges that there is such an order of priority, this order is only to guarantee that when, unfortunately, rationality and reasonableness conflict with each other, reasonableness should still subordinate rationality. In most of his writing we can still see that Rawls takes rationality seriously and even believes that a conception of justice must show that it is rationally justified. In his discussion of the question of stability, he argues that a satisfactory conception of justice must show that it is stable,⁸⁹ and stability depends on whether 'it is rational (as defined by the thin theory of the good) for those in a well-ordered society to affirm their sense of justice as regulative of their life'.⁹⁰ Rationality is an independent and crucial criterion in determining whether a conception of justice is satisfactory. Hence, a conception of justice must show that it is both rationally and reasonably justified. One would feel uncomfortable if the principles of justice always allowed for unreasonable treatments.

⁸⁶ Freeman (1990: 284).

⁸⁷ However, one might doubt that the source of normativity of Rawls's contractarian theory is not the dual conception of practical reason, but rather is the conceptions of person and society. I believe that these two conceptions have different functions. The conceptions of person and society define the context. Then rationality and reasonableness are applied in this context and explain the source of normativity. For a detailed explanation, see Note 63 in Chapter 2.

⁸⁸ Rawls (1980: 317).

⁸⁹ Rawls (1999a: 398).

⁹⁰ Rawls (1999a: 497).

But if the principles of justice often require one to sacrifice one's interests, then these principles should also be revised.

3.5.2 Hybrid contractarianism

Thus we can see that, according to Rawls, rationality and reasonableness are two separate criteria. Insofar as both rationality and reasonableness are independent conceptions of practical reason, and both of them are fundamentally important; a satisfactory contractarian theory should 'embody all the relevant requirements of practical reason'.⁹¹ Thus Rawls defines hypothetical contractors with two characteristics in order to model the dual nature of practical reason. Neither merely reasonable contractors nor merely rational contractors are adequate. Indeed, 'merely reasonable agents would have no ends of their own they wanted to advance by fair cooperation; merely rational agents lack a sense of justice and fail to recognize the independent validity of the claims of others'.⁹² Hence free and equal persons, who are the hypothetical contractors in Rawls' contractarianism, have two moral powers that correspond to the rationality and the reasonableness respectively.

By studying what free and equal persons would choose, we can know what political principles are both rationally and reasonably justified. The content of hypothetical agreement in hybrid contractarianism is different from that of Hobbesian and Kantian contractarianism. Since contractors are both rational and reasonable, they would consider both how to rationally advance their interests and how to behave reasonably to others. Their choice can specify the content of political principles that are justified from both rational and reasonable perspectives.

⁹¹ Rawls (1993: 90).

⁹² Rawls (1993: 52).

Chapter 3 Three Models of Contractarianism

[Principles of justice as fair terms of social cooperation] articulate an idea of reciprocity and mutuality: all who cooperate must benefit, or share in common burdens, in some appropriate fashion as judged by a suitable benchmark of comparison. This element in social cooperation I call the *Reasonable*. The other element corresponds to the *Rational*: It expresses a conception of each participant's rational advantage, what, as individuals, they are trying to advance. As we have seen, the rational is interpreted by the original position in reference to the desire of persons to realize and to exercise their moral powers and to secure the advancement of their conception of the good. Given a specification of the parties' highest-order interests, they are rational in their deliberations to the extent that sensible principles of rational choice guide their decision.⁹³

This hypothetical agreement has the character of the hypothetical agreement in Hobbesian contractarianism, for it represents a way to pursue the interests of contractors. But it also has the nature of the hypothetical agreement in Kantian contractarianism, for acting from it represents a way to be reasonable to others.

Further discussion of characteristics of hypothetical contractors and the content of hypothetical agreement will be presented in Chapter 6 and 7, where hybrid contractarianism is discussed in-depth. However, the discussion above is enough to show the special conception of practical reason of hybrid contractarianism, and how this element affects the other parts of this contractarian theory. In short, hybrid contractarianism can be seen as having three elements: (1) the dual conception of practical reason, (2) hypothetical contractors that are well-informed and both rational and reasonable, and (3) a hypothetical agreement that could be justified from both a rational and reasonable perspective. These three elements show that hybrid contractarianism has a distinctive content which is different from that of Hobbesian and Kantian contractarianism.

⁹³ Rawls (1980: 316).

3.6 Conclusion

Although contractarianism is a widely adopted methodology in justifying political principles, its classification is rarely discussed. A more precise conceptual classification can help us to capture the nature of some contractarian models more correctly. This chapter shows the inadequacy of the orthodox dichotomous model. In the last chapter I showed that the content of a contractarian theory fundamentally depends on its conception of practical reason. I then showed that there are three possible conceptions of practical reason. Correspondingly, there should be three models of contractarianism. If this is correct, then the orthodox dichotomy is flawed for it only divides contractarianism into two strands. The three models of contractarianism are presented in the following table.

Table 2 Three Models of Contractarianism

	Hobbesian contractarianism	Kantian contractarianism	Hybrid contractarianism
The conception of practical reason	The conception of rationality: reasonableness is subordinated to rationality	The conception of reasonableness: rationality is subordinated to reasonableness	The dual conception: rationality and reasonableness are co-equal to each other

The characteristics and circumstances of hypothetical contractors	Rational contractors who are well-informed utility-maximizers	Reasonable contractors who are well-informed and mutually respect one another	Well-informed contractors who are both rational and reasonable
The hypothetical agreement	Agreement that is mutually advantageous to everyone.	Agreement that could not be reasonably rejected by anyone	Agreement that could be justified from both rational and reasonable perspectives
Representative	Gauthier	Scanlon	Rawls

Finally, it is worth noting that these three models cannot be separated by clear boundaries. A contractarian can have a philosophical position between Hobbesian and hybrid models, or between hybrid and Kantian model. In fact, these three models should be viewed on a spectrum. Hobbesian and Kantian models are the two extremes, while the hybrid model holds a more central position. Different contractarians then occupy different points within this spectrum. Therefore, precisely speaking, all contractarians are hybrid contractarians, because no contractarians would occupy an extreme philosophical position such as ‘pure’ Hobbesian contractarianism, or ‘pure’ Kantian contractarianism. For example, even Gauthier, who is normally seen as the representative of Hobbesian contractarians, does not deny that people can honour the value of principles of justice because of its intrinsic moral value and not only because of its instrumental value in serving their interests.

In the last chapter of his *Morals by Agreement*, he argues that an individual will appreciate the intrinsic value of mutually advantageous principles which protected this society and hence will follow these principles even at the cost of his interests.⁹⁴ ‘Human beings have the capacity for an affective morality’ and hence are not merely economic persons who are only concerned with maximizing their utilities.⁹⁵

Scanlon, the representative of Kantian contractarianism, also argues that following moral principles can promote the well-being of followers. Although Scanlon does not say this point directly, this point is implicit in his theory of well-being. According to Scanlon, ‘what makes [a] thing good will...[be the fact that it] provides reasons for desiring it’.⁹⁶ For example, friendship is good because it can give us reason to pursue it. ‘[S]uccess in achieving [this good thing] becomes one of the things that make that person’s life better’.⁹⁷ Based on this theory of well-being, one can then recognise the good of following moral principles. For following moral principles implies that a person participates in a relation of mutual recognition. To Scanlon, following moral principles and protecting the relationship of mutual recognition are one and the same. On the other hand, the relationship of mutual recognition is a relationship that ‘is appealing in itself—worth seeking for its own sake’.⁹⁸ It means that this relationship itself can give us reason to pursue it, thus it should be seen as a ‘good’. Hence, by following moral principles, a person can maintain this relationship; by maintaining this relationship, a person’s life can be made better-off. Although Scanlon does not say that people will follow moral principles because this will promote their ‘good’, at least it seems that he does not think moral principles are completely irrelevant to a person’s good life.

⁹⁴ Gauthier (1986: 330-355).

⁹⁵ Gauthier (1986: 327).

⁹⁶ Scanlon (1998: 119).

⁹⁷ Scanlon (1998: 119).

⁹⁸ Scanlon (1998: 162).

We can hereby see that even the representatives of Hobbesian and Kantian contractarians are actually not as ‘pure’ as previously suggested. In fact, a contractarian usually has both Hobbesian and Kantian characteristics, but different contractarians have different proportions of these two elements. In this sense, all contractarians are hybrid and they differ only in proportion. Therefore, when I discuss ‘pure’ contractarian models such as the Hobbesian or Kantian, I will discuss the models in general and will only use the theories of these contractarians when they can be good examples of these models. For example, although Scanlon’s contractarian theory has alluded briefly to the good of following moral principles, this argument seems to be unimportant for he also says in a passage that his contractarianism is not ‘about what would be the most likely to advance [people’s] interests’.⁹⁹ Hence the Kantian element of his theory is far more substantial than its Hobbesian element; to the extent that the latter is even dispensable. I will focus on the Kantian part of Scanlon’s theory when I discuss Kantian contractarianism. These ‘impurities’ will be disregarded because my aim is to separately discuss the three models, Hobbesian, Kantian and hybrid contractarianisms, and show that all three models are problematic. Since these three models exhaust all sound ways of employing a contractarian methodology, the fact they are problematic will show that there are some critical flaws in the methodology. No matter how one employs this methodology, these flaws remain. Therefore, whether Gauthier or Scanlon are pure Hobbesian or Kantian contractarians is not a crucial issue, for, even if their theories are impure, their theories can shed light on the weaknesses of Hobbesian and Kantian contractarianism.

In the following chapters, I will discuss these three contractarian models. The normative force of these models depends on whether they can satisfy the condition of generality and the condition of priority. I will show that although Hobbesian and

⁹⁹ Scanlon (1998: 194).

Kantian contractarianism can satisfy the condition of generality, but cannot satisfy the condition of priority. The hybrid contractarianism is the only theory which takes the problem of priority seriously and tries to satisfy the condition of priority. However, this ambition hinders it from achieving a wide generality; therefore, hybrid contractarianism can only be applied to a small group of people. The condition of priority can only be satisfied at the cost of the condition of generality.

Chapter 4 Hobbesian contractarianism: Gauthier

4.1 Introduction

Rational interest always plays a crucial role in moral and political justification. As Kurt Baier observes, ‘throughout the history of philosophy, by far the most popular candidate for the position of the moral point of view has been rational interest’.¹ What better recommendation could there be for normative principles than the fact that they *pay* to be moral? It seems that if being moral or being just has no returns, then it is irrational to be so. From this we can see the considerations of rationality are common in our practical reasoning. As we saw in the last chapter, rationality is a weighting model that ranks our preferences in terms of utility and motivates us to maximize our individual utilities. This instrumental conception of rationality is widely endorsed in various subjects, and Hobbesian contractarians take it to be their conception of practical reason. In this chapter, I will further discuss how Hobbesian contractarians justify political principles by hypothetical agreement based on this assumption,² and why, because of this assumption, Hobbesian contractarianism fails to fulfill the condition of priority.

Before we start to discuss Hobbesian contractarianism, I should first clarify the relationship between Hobbes and Hobbesian contractarianism. Hobbesian contractarians normally claim that their roots are in Hobbes (and some argue that Hume is also in this tradition).³ In general, this is not a wholly false claim because

¹ Baier (1958: 187).

² Some people might argue that Gauthier’s contractarian project is more concerned with moral justification. Yet Gauthier also acknowledges that his contractarian theory has a political dimension and can be used to justify social institution and practice. See Gauthier (1997).

³ Gauthier (1979).

they are profoundly influenced by Hobbes in at least two ways. First, their view of human rationality is affected by Hobbes. Hobbes believes that human beings are creatures of passions and desires. Human life, then, is a journey of seeking to satisfy desires: 'Life itself...can never be without desire'.⁴ Hobbes sees human beings as 'maximizers' who aim at maximizing the satisfaction of desires. This maximizing behaviour is 'a general inclination of all mankind, a perpetual and restless desire of power after power, that ceaseth only in death'.⁵ A similar understanding of human nature can also be seen in the writings of Hobbesian contractarians, who believe that Hobbes' words can be articulated better by the theory of rational choice.⁶ Secondly, the way that Hobbesian contractarians understand how people form a social agreement is also highly similar to Hobbes'. According to Hobbes, since human beings are rational, they should understand that they can only get rid of the state of nature by subjecting themselves to an absolute sovereign. Therefore, a social agreement which authorizes an absolute sovereign is justified, because it is a mutually advantageous agreement which avoids the poor state of nature and brings benefits to all people.⁷ Although, these days no Hobbesian contractarians would agree with absolutism, none of them would deny that a justified social agreement should be a mutually advantageous agreement.⁸ Because of these two similarities, Hobbesian contractarians claim that their approach can be traced back to Hobbes and they are indebted to his insights.

However, there are also some critics who argue that Hobbesian contractarianism misunderstands Hobbes.⁹ The difference between Hobbes and the 'so-called' Hobbesian contractarians should not be overlooked. For example, some people argue

⁴ Hobbes (1994: 35-36).

⁵ Hobbes (1994: 58).

⁶ See Gauthier (1986: 21-23), Hampton (1986: 16-17) and Kavka (1986: 38-40).

⁷ Hobbes (1994: 79-81).

⁸ See Gauthier (1986: 14), Hampton (1986: 138-147) and Kavka (1986: xii-xiii).

⁹ Ryan (1988: 92).

that Hobbes is concerned only with the desire for self-preservation, and they justify the hypothetical contract by the satisfaction of this primary desire. However, Hobbesian contractarians are not concerned specifically with this preference, and they justify the hypothetical contract by the maximization of individual utility, which consists of other preferences apart from self-preservation. Whether 'Hobbesian contractarianism' is a correct label could be the subject of a chapter of its own, but I would like to bracket this question here first. Since my interest is in the normativity of Hobbesian contractarianism, regardless of whether or not Hobbesian contractarians misunderstand Hobbes, this theory itself is well-structured and worth discussing. Hence I will leave this question aside for the moment and simply keep using the term 'Hobbesian contractarianism'.

As I said in Chapter 2, the normativity of a hypothetical contract depends on whether it can satisfy the condition of generality and the condition of priority. The purpose of this chapter is to examine whether Hobbesian contractarianism can satisfy these two conditions. I will use David Gauthier as the representative of Hobbesian contractarianism. First, I will explain how the conception of rationality determines the characteristics of hypothetical contractors and the content of hypothetical agreement, which are the other two elements in Hobbesian contractarianism. Based on the conception of rationality, Hobbesian contractarians argue that a political principle is justified only when it is mutually beneficial, since such a principle represents an agreement that is rationally justified. Secondly, I will discuss the inadequacy of this assumption. This assumption fails to fully model our practical reason because it overlooks another aspect of practical reason—reasonableness. This flaw explains why although Hobbesian contractarianism can satisfy the condition of generality, it still fails to satisfy the condition of priority. The third section will be a discussion of possible objections from Hobbesian contractarians. Hobbesian contractarians might respond that reasonableness is in fact not independent since it can be reduced to one of the preferences in a person's utility function. I will argue that this reduction is unsound because it distorts the very nature of reasonableness.

4.2 Hobbesian contractarianism: an elaboration

Hobbesian contractarians believe that actual people deliberate about practical issues in terms of rationality. If political principles are rationally justified, then actual people will have reason to accept them. These political principles can be specified by the hypothetical agreement that would be made by rational hypothetical contractors. In this section, I will illustrate how these contractors are defined and how decisions are arrived at.

4.2.1 Characteristics of hypothetical contractors

According to my definition in Chapter 3, rationality is a decision-making process which takes every motivation of people as preferences, ranks different preferences in terms of utility and aims at maximizing the utility of a person. Hobbesian contractarians, like Gauthier, believe that if the hypothetical contractors can deliberate in accordance with rationality, then the decision of hypothetical contractors is justified to actual people. Actual people can identify themselves with the hypothetical contractors and understand the rationale behind the decisions of contractors. Based on this belief, the characteristics of hypothetical contractors are defined in order to represent a pure rational process of deliberation. These characteristics can be shown in three aspects: the motivation of contractors, the qualification of contractors and the knowledge of contractors.

The motivation of contractors

Hypothetical contractors are defined as purely rational contractors who have *coherent* and *considered* preferences.¹⁰ These rational contractors have different sets of preferences, which constitute their own utility functions. Based on their utility function, rational contractors aim at nothing except maximizing their utility effectively: ‘The rationality is assumed and is identified with the aim of utility-maximization’.¹¹ Moreover, contractors are assumed to be mutually unconcerned with one another (take no interests in one another’s interests).¹² Each rational utility-maximizer has their own utility function which is independent of those of others. This assumption easily attracts a critique that it is counterfactual, for people ‘do, at least to some extent, care about how other people’s conceptions of the good are promoted’.¹³ But Gauthier does not intend to make such an absurd assumption. His assumption is more restrictive than many critics think: he requires only that people are not interested in the utility of those with whom they exchange. ‘This is Wicksted’s requirement of “non-tuism”; my preferences do not involve you, although they may involve some third person not party to our interaction’.¹⁴ That is to say,

¹⁰ As I discussed in Chapter 3, these two conditions are the necessary conditions for a preference to be counted in a utility function. These two conditions prevent those inconsistent and ill-considered preferences from affecting the bargain. But how to judge whether a preference is considered or not? Some people object that Gauthier cannot avoid assuming some *a priori* normative commitments and obligations in order to show that some preferences are ill-considered (such as the preference to mistreat others) while some are not. Since this critique is not highly relevant, I will discuss a possible response to this critique only briefly here. I do not think Gauthier is vulnerable to this critique, for he could simply claim that he only requires rational contractors to think about their preferences carefully and he trusts that, in general, those ill-considered preferences, such as the preference to mistreat others, could be excluded after reflection, but this reflection does not necessarily involve normative commitments or obligations. For example, as James Griffin suggests, a desire could be seen as ‘informed desire’ insofar as people ‘appreciated the true nature’ of the object of these desires, and this appreciation could be irrelevant to normative commitments. (See Griffin 1986: 11) For an example of this critique, see Brandom (2001), Ripstein (2001), Baier (1988) Scanlon (1975: 78).

¹¹ Gauthier (1975: 209).

¹² Gauthier (1986: 10-11, 102-103).

¹³ Vallentyne (1991c: 5).

¹⁴ Gauthier (1986: 87).

rational utility-maximizers are not *wholly* mutually unconcerned with one another. Sometimes the utility function of contractors involves the utility function of others, but in most cases, the utility functions of contractors are independent of one another.

The qualification of contractors

In Gauthier's hypothetical contract, only those rational utility-maximizers whose cooperation would benefit others qualify. Those people who are not able to offer benefits, such as children, the severely handicapped and members of future generations are not included as contractors.¹⁵ This scope can be explained by the conception of rationality. Since the hypothetical contractors aim only at rationally maximizing their utility, it would be irrational for them to cooperate with people who could not bring them a gain in utility. The only form of cooperation to which they will agree is a mutually advantageous agreement. Consequentially, those 'powerless' can be included in the scope of contract only when the rational utility-maximizers, who can participate in the contract, care about them personally. Nevertheless, even if those 'powerless' were loved, their moral standing would still be *secondary*, since it would depend only on the personal care of rational utility-maximizers, who have the *primary* moral standing.¹⁶

The knowledge of contractors

Rawls' contract is famous for the veil of ignorance imposed on contractors. Contractors only have general knowledge of their society but know nothing about their particular features. Gauthier consciously distances himself from this approach

¹⁵ Gauthier (1986: 268-269).

¹⁶ Morris (1991: 76-95).

and argues that his contract is based on rational negotiations among well-informed people.¹⁷ In Gauthier's contract, contractors are fully informed about their capacities and social situations. The purpose of making this assumption is clearly to reduce the gap between actual people and hypothetical contractors. If hypothetical contractors were ignorant of particular information, then they would have a utility function which is highly different from that of actual people. The decision that hypothetical contractors would make in order to maximize their utility would be rationally unacceptable to actual people. Hence, in order to ensure that the rational decisions of actual people and hypothetical contractors are the same, Gauthier rejects any veil of ignorance in his contractarian theory.¹⁸

4.2.2 The state of nature: the Prisoners' Dilemma

These three characteristics constitute the hypothetical contractors in Gauthier's contractarian theory—well-informed rational utility-maximizers who are mutually unconcerned with one another and cooperate with others only when others can benefit them. Gauthier believes that the decisions of these hypothetical contractors represent what are rationally justified to actual people. Hobbesian contractarians invite us to imagine a 'state of nature', that is, a situation that these hypothetical contractors were grouped together. In particular circumstances, these contractors can

¹⁷ Gauthier (1986: 5)

¹⁸ But there is another critique that well-informed reasoning is itself unrealistic. Someone might argue that the utility function of actual people usually involves wrong information or unconsidered preferences. Hence the utility function of well-informed hypothetical contractors is still different from that of actual people. The actual people may still fail to identify themselves with hypothetical contractors. It is true that wrong information and unconsidered preferences inevitably exist in our real life. Nevertheless, actual people rarely think that these factors are desirable and relevant. Even if well-informed maximization is impossible in the real world, it is still *instructive* for actual people. No one would prefer pursuing a goal because of wrong or inadequate information. More precisely, utility-maximization is valuable only when people are fully-informed, for people usually would not value utility-maximization when the utility function is affected by fraud. It is commonly acceptable that rational utility-maximization should not be contaminated by these factors. For how the decision of hypothetical contractors act as an instruction of actual people, see Morton (2001)

spontaneously achieve a Pareto-optimal outcome: an outcome where no individual could be made better-off without some individuals being worse-off, without any agreement. Gauthier believes that, under the conditions of a perfect market, the divergent and seemingly opposed interests of different individuals fully harmonize with one another. All contractors can do as well as they can, that is, their individual utility can be maximized subject to the utility maximization of others. Hence, no rules or agreement are required in the perfect market, because a mutually beneficial order can be spontaneously achieved by the interactions among contractors. ‘The perfect market, were it realized, would constitute a morally free zone, a zone within which the constraints of morality would have no place. In leaving each person free to pursue her own interest in her own way, the market satisfies the ideal of *moral anarchy*’.¹⁹

However, this perfect outcome does not always occur. In some cases, market failures appear and the rational pursuit of individual utilities leads to suboptimal outcomes, that is, outcomes where one individual could be made better-off without any individuals being worse-off. One of the best-known examples is the ‘Prisoner’s Dilemma’. It describes a situation where two prisoners are captured and interrogated in separate rooms. Each is told that, if they confess before the other does, then they will only be sentenced to one year’s imprisonment while the other will be sentenced to ten years. If neither confesses, then both will only be sentenced to three years. But if both confess, then both will be sentenced to eight years. Their situation can be represented by the following matrix:

Table 3 Prisoners’ Dilemma

¹⁹ Gauthier (1986: 84).

A's strategy \ B's strategy	Confess	Not confess
Confess	-8, -8	-1, -10
Not confess	-10, -1	-3, -3

In this matrix, the left-upper cell, the convergence of the rational choice of both prisoners, is a suboptimal outcome, because there is an alternative outcome (the right-lower cell) that both prisoners could gain more benefit without making other worse off. Although each prisoner knows that the optimal outcome is the right-lower cell, each will ironically fall into the suboptimal left-upper cell, because each can maximize his utility only by confessing, irrespective of whether or not another prisoner confesses. We can see the divergence between rationality and optimality: even all people act rationally, they cannot achieve the optimal outcome.

This is why a social contract is needed. In the Prisoner's Dilemma, both prisoners will choose to confess only because they do not share the appropriate understanding. However, if they reached an agreement before they were caught, both of them would choose to not confess and hence they can achieve the optimal outcome. Gauthier believes that those contractors can get rid of market failures in a similar way. Market failures appear because contractors act independently without cooperating with each other. So what if they make a social agreement instead? Contractors can make an agreement with one another which guarantees that every contractor will choose a

collective strategy that leads them to an optimal outcome and provides them the opportunity to avoid market failure.

4.2.3 The hypothetical contract: the Lockean proviso and the principle of minimax relative concession

For the sake of leaving the suboptimal position, these contractors bargain with each other and form a social agreement which is mutually advantageous. However, bargaining needs a baseline which can determine the goods that each person brings to the bargaining table and that is not subject to bargaining. Without this baseline, one cannot understand the idea of advantage.²⁰ Hence, before bargaining, contractors have to know what goods they can legitimately own first. Gauthier suggests that this 'initial distribution of goods' is defined by a proviso on previous acquisition of resources. Gauthier, following Nozick, calls this proviso the 'Lockean proviso'.²¹ Goods belong to a person when they are 'acquired by him without taking advantage of any other person—or, more precisely, any other co-operator'.²² Here 'taking advantage' of another person refers to making the situation of another person worse in order to better one's own. The function of this proviso is to exclude irrelevant factors such as coercion and free-riding. Under the constraint of this proviso, contractors can no longer acquire goods by coercive or predatory activities. They can get the goods only by their own labour. Since the Lockean proviso guarantees a fair initial situation without coercion, the hypothetical agreement which is reached under the constraint of this proviso will be a fair agreement.

²⁰ Barry (1989: 12-15).

²¹ Nozick (1974: 175-182).

²² Gauthier (1986: 201).

At the bargaining table, contractors have to think about how the cooperative surplus should be divided. They know that by cooperating with each other, extra goods will be created. The question is how these goods should be divided among contractors who create them together. According to Gauthier, since society's members are rational utility-maximizers, they will seek to get as much cooperative surplus as possible during bargaining. If the claims of these society members are incompatible, there is a second stage in which each member offers concessions to the others by withdrawing some portion of their original claim and proposing an alternative outcome. Concession making continues until a set of mutually compatible claims are reached. According to Gauthier, this is reached when each member makes concessions that are relatively equal to the concessions of others. Here the relative concession refers to the concession that one has to suffer when one bargains with others in the division of cooperative surplus:

The relative concession that a person makes for a given option is the ratio of (a) the excess of (i) the utility for that person of his/her most favorable admissible option over (ii) the utility for that person of [the agreement] to (b) the excess of (i) the utility for that person of his/her most favorable admissible option over (ii) the utility for that person of the initial bargaining position option. An admissible option is one that is both feasible and accords everyone at least as much utility as the initial bargaining position.²³

Gauthier argues that the outcome of bargaining will be the 'principle of minimax relative concession', which is a bargaining solution that minimizes the maximum relative concession that anyone makes. This principle is most rational because it 'is most favourable to [each society member], minimizing the costs of her restraint and maximizing the benefits she receives from the restraint of others'.²⁴ In every other alternative outcome, the maximum relative concession would be greater, which

²³ Vallentyne (1991c: 8).

²⁴ Gauthier (1986: 128).

means that some society members would be disadvantaged and would be required to sacrifice more. As Vallentyne says, 'the intuitive idea behind this solution is that, since one's ground of complaint can be measured by one's relative concession, minimizing maximum relative concession minimizes the grounds for complaint'.²⁵ Therefore, the principle of minimax relative concession is rationally the most acceptable solution to everyone, because any other alternative solution would impose more burdens on some members and would be rejected by these members.

The principle of minimax relative concession is the outcome of a joint strategy choice.²⁶ By following the principle of minimax relative concession, hypothetical contractors can overcome market failure and achieve an optimal outcome. However, someone might doubt whether, even if it is rational to make such an agreement, it is really rational to comply with it. It seems that, if individuals are by nature rational utility-maximizers, they should free-ride on others and break the agreement given that others must still honour the agreement. Gauthier disagrees that this is rational. He distinguishes two strategies for maximizing utilities. One is straightforward maximization, which means that an individual 'seeks to maximize his utility given the strategies of those with whom he interacts'.²⁷ Without concern about any other things. Another is constrained maximization, which means that an individual 'has a conditional disposition to base her actions on a joint strategy, without considering whether some individual strategy would yield her greater expected utility...In other words, a constrained maximizer is ready to cooperate in ways that, if followed by all, would yield outcomes that she would find beneficial and not unfair'.²⁸ Gauthier argues that provided the government can guarantee a sufficient degree of

²⁵ Vallentyne (1991c: 8-9).

²⁶ Some theorists argue that rational contractors might choose other principles apart from the principle of minimax relative concession. This principle is only one of the possible joint choices. For a discussion of this question, see Hardin (1988).

²⁷ Gauthier (1986: 167).

²⁸ Gauthier (1986: 167). For the discussion of constrained maximization, see also Gauthier (1975)

translucency, that is, other people can have a fairly good idea what we are really like and free-riding is difficult, constrained maximization is a more rational strategy than straightforward maximization. Individuals should be disposed to act on the basis of the agreement, rather than maximizing their utilities directly and free-riding on others.

This concludes the exposition of Hobbesian contractarianism. We can see how a Hobbesian contractarian develops a whole theory from the conception of rationality. In Gauthier's contract theory, hypothetical contractors are well-informed rational utility-maximizers, who model the rational deliberative process of actual people. The main problem for these contractors is market failure, which is the suboptimal situation that contractors would fall into if they maximized their utility without any agreement. So they have to make an agreement with one another in order to get rid of these suboptimal situations and enjoy a cooperative surplus. Before bargaining, contractors have to define what goods belong to them, so they first agree on the Lockean proviso which defines the initial distribution of goods. This initial distribution determines the bargaining power of contractors. Based on their bargaining power, contractors bargain with one another and at last agree on the principle of minimax relative concession. This principle acts as a public constraint which helps them to reach optimal, mutually beneficial outcomes. Contractors will also change into constrained maximizers and be disposed to maximize their utility under the public constraint. Hence, Gauthier shows how substantial political principles are derived from a formal conception of rationality by the contractarian method.

4.3 Hobbesian contractarianism and the condition of generality

One of the crucial conditions of having normativity is the condition of generality. Contractarians must show that the social agreement is not only justified to the

hypothetical agents, but also justified to actual people. So the conception of practical reason should be generally acceptable to actual people. The strength of Hobbesian contractarianism is that it assumes a weak conception of rationality. This section will show how Hobbesian contractarianism can fulfill the condition of generality.

4.3.1 The general acceptability of the conception of rationality

Gauthier's solution to the problem of generality is to rely on the weak conception of rationality, which is a conception 'without any prior moral assumptions'.²⁹ In fact, this strategy of avoiding moral assumptions is also the common strategy of Hobbesian contractarianism since making moral assumptions would lead to controversy.³⁰ People who do not care so much about impartiality might refuse to identify themselves with the hypothetical contractors who have a strong motivation to be impartial. The relationship between hypothetical contractors and actual people would become unclear. In order to prevent this controversy, Hobbesian contractarians propose that a contractarian theory should appeal to 'a non-moral, or morally neutral, base'.³¹

Hence, Hobbesian contractarians, including Gauthier, borrow the conception of rationality from the theory of rational choice, which does not have any prior moral presupposition and provides only a 'weak and widely accepted conception of practical rationality'.³² Gauthier believes that maximizing individual utility is the fundamental characteristic of human beings.³³ It is hard to prove the generality of the conception of rationality; not because it is groundless, but rather because its

²⁹ Gauthier (1986: 6).

³⁰ Kavka (1986: 64-65) and Hampton (1986: 22).

³¹ Gauthier (1986: 17).

³² Gauthier (1986: 17).

³³ Gauthier (1977: 332).

generality seems to be self-evident. Yet I can give two reasons to support this claim: the first is empirical; the second is theoretical. First, the generality can be proved by current neuroscientific studies. According to some neuroscientists, the utility calculation model is actually a decision-making structure that physically exists in our brain.³⁴ This structure quantitatively encodes the relative desirabilities of all possible courses of action. Our brain then chooses among these courses of action according to the relative desirabilities. This mechanism is just like the rational decision-making process that we discussed before. Based on these empirical studies, neuroscientists argue that the utility calculation model is not merely an arbitrary assumption that a person could choose to accept or not, but rather an innate neurobiological structure of the brain each person is born with. This empirical study can also explain why the utility-function decision-making model is transcultural; it persists in many societies in various forms.³⁵

We also have a second reason to accept this conception of rationality. As Binmore argues, the conception of rationality does not specify what goals people must have. These rules simply require people to be consistent in pursuing their goals.³⁶ Regardless of the goals of people, they will be hard pressed to deny that they have to pursue these goals consistently. Even though sometimes people might behave irrationally and fail to maximize their individual utility, it does not mean that rationality does not occupy a crucial role in their practical reason. Hence it is sensible

³⁴ See Glimcher, Dorris and Bayer (2005). For a psychological point of view, which argues that people acquire this model of decision-making in an *a posteriori* way, through living in a culture and learning early in life, see Morton (2001).

³⁵ See Miller (1999) and Rapoport and Chammah (1965).

³⁶ Binmore (1994: 21, 27).

to claim that, compared with other interpretations of practical reason, rationality is a general interpretation which is acceptable to most of the people.³⁷

These two reasons can also shed light on the question of why so many contractarians adopt the Hobbesian approach. The standard critique of contractarianism is that the relationship between the hypothetical contract theory and the real world is unclear. However, Hobbesian contractarianism can avoid this problem for it is based on a weak conception of rationality borrowed from the theory of rational choice. It is a minimal assumption that does not invoke any controversial moral assumptions and assumes only that individuals are rational utility-maximizers. In fact, this is also the distinctive strength emphasized by Hobbesian contractarians. As Kavka points out, unlike philosophers who imagine human beings as too 'good', 'Hobbesians...see a system of plausible moral and political hypotheses emerging from a realistic portrayal of human nature'.³⁸

By grounding their theory on a widely accepted assumption, Hobbesian contractarians believe that actual people will have no problem in identifying themselves as hypothetical contractors. The hypothetical contractors are rational persons who aim at maximizing utility effectively, but they are also familiar with the 'capacities, situations, and concerns' of actual people, hence their decisions should also be able to promote the utility of actual people.³⁹ The contract that they will

³⁷ Here the claim that rationality is generally acceptable merely means that people are bound by the rules of rationality in their practical reasoning and they would usually agree that they have an obligation to follow these rules. But it does not mean that people generally accept rationality as the *only* component of practical reason. As we will see later in the thesis, the Hobbesian conception of practical reason overlooks the reasonable aspect of practical reason. Apart from rational utility-maximizers, people also generally conceive themselves as reasonable moral persons who are concerned with justifiability. However, it does not affect the general acceptability of the conception of rationality.

³⁸ Kavka (1986: 80).

³⁹ Gauthier (1986: 9).

make is a mutually advantageous agreement which can bring benefits to actual people. The actual people, who also want to maximize their utilities rationally, will not find this agreement weird and will accept this agreement. Therefore, the Hobbesian contract is not an authoritarian scheme which imposes arbitrary restrictions on people by relying on an irrelevant contract story. Rather, the Hobbesian contract explains what political principles are rationally justified and actual people who are rational should comply with these political principles.

4.3.2 Presupposing hedonism?

However, some critics might argue that the conception of rationality is less acceptable than those rational choice theorists believe. The wide acceptance of this conception of rationality depends on its neutrality to all goals. But critics could argue that there is actually a hedonistic goal behind the conception of rationality. The neutrality of this conception is merely an illusion for it presupposes a substantial, ultimate goal that all human beings should pursue: happiness.⁴⁰ Rationality presupposes that all preferences are commensurable. This implies that it must presuppose an ultimate preference above these preferences, and the weight of each reason is measured by how much it can contribute to the satisfaction of this preference. These critics then argue that the ultimate preference should be the maximization of happiness, which is measured by some subjectively identifiable psychological intensity.⁴¹ Hence how we should arrange the importance of other preferences depends on how much each preference can contribute to the amount of our happiness. Since the maximization of happiness, which is a particular psychological quality, is itself a substantial, distinctive conception of the good, critics argue that the conception of rationality is in fact not neutral. This conception

⁴⁰ Korsgaard (1999).

⁴¹ Korsgaard (1999: 97-98).

‘cannot after all be detached from its origins: it is a child of introspective psychology, grounded in the British empiricist theory of happiness or the good’.⁴² If the conception of rationality presupposes a hedonistic goal, then this conception might fail to be an impartial process with respect to all preferences. This casts doubt on how general the conception of rationality is.

Nevertheless, this critique is actually based on a misunderstanding of the nature of utility. It wrongly assumes that utility represents a measurable quantity of happiness, thus utility-maximization should presuppose a goal of maximizing happiness. However, utility does not refer to any specific psychological quality. Rather, it is simply a measure of preferences which are ranked in accordance with principles of coherence, such as the principle of transitivity. In fact, as some rational choice theorists emphasize, the most crucial idea of utility maximization is coherency rather than happiness.⁴³ Hobbesian contractarians do not assume that people rank their preferences according to how much these preferences contribute to their happiness. People can rank preferences according to other grounds, such as how much these preferences can contribute to the promotion of certain communal values. What ground people should choose is up to them, and the answer is not necessarily happiness. Hobbesian contractarians are concerned only about whether the preferences are ranked in a consistent order, regardless of whether these preferences contribute to their happiness. If a preference A has a higher utility to a person than a preference B, then it does not mean that A could bring this person more of particular kind of psychological experience than B could offer. Rather, it means that this person prefers A to B only, and it has nothing to say about the psychological state of this person. Even if A would undermine the happiness of a person, A still has a higher utility.

⁴² Korsgaard (1999: 98).

⁴³ See Luce and Raiffa (1957: 31-32), Arrow (1967: 5), Harsanyi (1986: 8) and Debreu (1959: 55-59).

Therefore, maximizing utility neither presupposes a calculation of the quality of a certain kind of psychological experience, nor presupposes that all our actions serve an ultimate purpose of maximizing this psychological experience. It presupposes only that a person can have a consistent order of preferences. That is why Gauthier says,

We have no reason to believe that preference has a unique, measurable object such as [quality of experience or enjoyment]. It seems clear that preferences cannot be correlated with any single dimension or characteristics of the state of affairs that they relate. Indeed, it seems clear that preferences do not depend solely on the qualities of experience.⁴⁴

4.3.3 Presupposing egoism?

Critics can further answer this response by saying that, even if rationality does not presuppose that people have a hedonistic goal of pursuing their happiness, it still presupposes that people have an egoistic goal of pursuing their self-interest. The utility function only takes preferences which are related to a person's self-interest into account. For example, Geoffrey Sayre-McCord objects that the assumption of Hobbesian contractarianism is counter-intuitive because human beings are not wholly self-interested. They have altruistic preferences which are not determined by selfish or narrowly self-interested desires.⁴⁵

Again, this critique is still based on a misunderstanding of the nature of utility. Rationality does not presuppose that people are egoistic. Since utility is only a neutral idea, it says nothing about whether preferences in a utility function are

⁴⁴ Gauthier (1986: 36).

⁴⁵ Geoffrey Sayre-McCord (1991) believes that Hobbesian contractarianism is based on an assumption of self-interestedness. See also Morris (1988).

necessarily related to a person's self-interest. People can have an altruistic preference, that is, they prefer a state of affairs in which other people are benefited but their own self-interest is harmed. In this case, their utility can still be maximized after their altruistic preference is satisfied. Rational calculation does not exclude this kind of preference.⁴⁶ Regardless of whether a person's preferences are self-interested or altruistic, given that these preferences are ranked consistently, they are counted as legitimate preferences and should be taken into account.⁴⁷ That is why Gauthier says that Hobbesian contractarianism does not presuppose a substantial egoistic conception of the person:

Neither conception of rationality requires that practical reasons be self-interested. On the maximizing conception it is not interests in the self, that take oneself as object, but interests of the self, held by oneself as subject, that provide the basis for rational choice and action. On the universalistic conception it is not interests in anyone, that take any person as object, but interests of anyone, held by some person as subject, that provide the basis for rational choice and action.⁴⁸

4.3.4 Too unrealistic?

Apart from the criticisms of presupposing hedonism and egoism, the conception of rationality could also be condemned as being unrealistic. Some critics may argue that the axioms that are taken to characterize rationality are matters of controversy because, from empirical experiments, actual people's choices are always the reverse of what the model of rationality would predict. For example, the conception of rationality requires the order of preferences to be complete, which means that 'for

⁴⁶ Some theorists have pointed out the model of utility maximization does not presuppose that all desires of people are self-interested. See Arrow (1963: 3) and Sen (1977: 322-324).

⁴⁷ Gauthier (1986: 23).

⁴⁸ Gauthier (1986: 7).

any two possible outcomes in a choice situation, the chooser must either prefer one to the other or be indifferent between them...[The principle of completeness rules out preferentially non-comparable outcomes'.⁴⁹ However, as some economists argue, if we take the principle of completeness seriously, then it is actually a very demanding principle.⁵⁰ Both Herbert Simon and George Shackle argue that future situations and consequences are incalculable.⁵¹ The world is full of endogenous factors that lead to complexity and uncertainty. How can I be sure that my choice can lead to the predicated outcome? How can I be sure that my choice must be better than any other alternatives if I look back one day in the future? Due to the complexity and uncertainty, it is impossible for people to accurately predict the utility of each preference and compare among various preferences. Any real future is inherently and fundamentally unknowable. Hence, a complete order of preferences is in fact an unachievable task for actual people.

Similarly, the principle of transitivity required by rationality is also vulnerable to a similar critique. Transitivity requires rational people not to rank their preferences as a cycle, that is, they cannot prefer A to B, prefer B to C, then prefer C to A, for '[s]uch cycle would divest choices of all rationality'.⁵² Rather, rational people should rank their preferences in a consistent order. For example, if they prefer A to B and B to C, then they should prefer A to C. This principle is also criticized as being too harsh because empirical experiments show that most of the people do not choose in this way. In a famous betting experiment conducted by Maurice Allais, Allais shows that people may make intransitive choices in different gambles.⁵³ For example, in the Table 4, people may prefer A to B in one gamble because B is a large-stake small-prize gamble. They may also prefer B to C in another gamble because C is a large-

⁴⁹ Gauthier (1986: 39).

⁵⁰ I am indebted to Albert Weale for this point.

⁵¹ Simon (1962) and Shackle (1964)

⁵² Gauthier (1986: 41).

⁵³ Allais (1979)

stake small-prize gamble. But if they have to choose between A and C in one gamble, then they may prefer C to A because now C looks like a small-stake large-prize gamble.⁵⁴ Hence, the principle of transitivity is, like the principle of completeness, controversial because it is always inconsistent with the actions of actual people.

Table 4 A version of Allais Paradox

<div> <div>Probability</div> <div>Choice</div> </div>	Probability 0.4	Probability 0.2	Probability 0.4
A	6	6	6
B	0	10	10
C	0	0	15

Although these two critiques are valid, they do not affect my claim because they are challenges to the *predictive* power of rationality, but not the challenges to the *normative* power of rationality. As a model of practical reasoning, the conception of rationality has two functions. One is to predict what actual people would do under

⁵⁴ This example and table is borrowed from Loomes and Sugden (1982: 815-816), which is a paper about Allais paradox.

certain circumstances; another is to prescribe what actual people ought to do as a rational being. The critiques of Simon, Shackle and Allais correctly point out that the predictive power of the conception of rationality is dissatisfactory. However, what I would like to argue is that the normative requirements of the conception of rationality are generally accepted by actual people. People may act irrationally in their daily lives because of laziness, prejudices, wrong information or inadequate information. But when they were told a more rational suggestion, most of them would prefer this suggestion and would even think that they *ought to* accept this suggestion.

For example, suppose that you have a holiday but you cannot decide whether you should go to Paris or Rome. Then you know that there is a super computer which can accurately predict the outcomes of your choices. The computer tells you that you would have an unhappy time in Paris, thus it suggests that you should go to Rome instead. Given that the calculation of the super computer is reliable, I believe that most of the people would accept its suggestion. This shows that even if people are incapable of having a complete order of preferences in their daily lives, they would prefer having a complete order and being more rational. Few people would reject the normative requirement of being rational. Therefore, even though the conception of rationality fails to predict human behavior accurately, it does not affect the generality of Hobbesian contractarianism. The condition of generality is concerned with whether actual people would generally accept what the conception of rationality prescribes, but not whether actual people would act as what the conception of rationality predicts.

In this section I illustrated the structure of Hobbesian contractarianism and how it satisfies the condition of generality. Hobbesian contractarians understand that the practical reason of human beings is governed by the conception of rationality. Based on this conception of practical reason, Hobbesian contractarians develop a theory to

show that rational contractors will agree on political principles which are mutually advantageous. This contractarian theory satisfies the condition of generality because actual people, similar to hypothetical contractors, generally accept the normative requirements of the conception of rationality. So the decisions of hypothetical contractors, which are made in accordance with these requirements, should also be justified to actual people. Some critics wrongly think that this instrumental conception of rationality is not neutral because it takes hedonism and egoism as the underlying assumptions. However, these critiques are merely based on a misunderstanding of the formal nature of utility. Since the rational calculation also includes non-egoistic and non-hedonistic preferences, accepting this conception of rationality does not imply endorsing hedonism or egoism. Also, although the conception of rationality sometime wrongly predicts the behaviours of actual people, the normative requirements of the conception of rationality are still generally accepted. Nevertheless, in the remainder of this chapter, I shall argue that, although Hobbesian contractarianism can satisfy the condition of generality, it fails to satisfy the condition of priority because it overlooks another aspect of practical reason.

4.4 Reasonable rejectability of Hobbesian contractarianism

In Chapter 2, I argued that a satisfactory hypothetical contract not only has to fulfill the condition of generality, but also has to fulfill the condition of priority. Contractarians have to show that the hypothetical contractors have already taken all the rules that govern practical reason into account, so that they can completely model the practical reasoning of actual people. Hobbesian contractarianism is an appealing approach of contractarianism because it can model the rational aspect of our practical reason. However, I shall argue that this approach is inadequate because it fails to model the reasonable aspect of practical reason. The consequence of failing to model reasonableness is that Hobbesian contractarianism is unjustified from the *reasonable* perspective.

4.4.1 The reasonable aspect of practical reason

Practical reason has two aspects: rationality and reasonableness. Hobbesian contractarianism captures the former but overlooks the latter. The conception of reasonableness is a ‘testing’ model which has an absolute moral criterion: the constraint of justifiability. It uses this constraint to evaluate desires. If a desire motivates people to violate the constraint of justifiability, then no matter how important this desire may be, it will still be rejected. The constraint of justifiability is a constraint that requires people to behave in a way which can be justified to others. Reasonableness presupposes that people share an aim of living with others who are also reasonable.⁵⁵ Because of this aim, people acknowledge that they have to respect the capacity for reason-giving of others. All people are capable of giving justifications to defend their actions, assessing justifications of others, and asking for justification from others. Hence reasonableness requires that people should respect the claims of others and should ensure that their actions are justifiable to others. If others then suggest a strong reason to reject their action, they should refrain. This respect is different from the respect required by rationality. For rationality merely requires people to respect others on condition that this behaviour could maximize their own utilities, whereas reasonableness requires people to respect others *unconditionally*, no matter how their utilities would be affected.

I have briefly discussed how reasonableness is a generally accepted conception of practical reason in Section 3.3.2, and will further elaborate it in Section 5.2.4. Hence, apart from rationality, which simply motivates people to pursue interests rationally, people are also reasonable in as much as they generally have a motivation to behave in a way which could be justifiable to others. From a reasonable perspective, the

⁵⁵ Scanlon (1998: 154).

pursuit of interests is an irrelevant consideration. If some actions are prohibited by the constraint of justifiability, then people will still refrain from doing these actions even at the cost of their own interests.

4.4.2 Why Hobbesian contractarianism is to be reasonably rejected? (1): the powerlessness critique

The reasonable aspect of practical reason can explain why Hobbesian contractarianism is objectionable even though it is rationally justified. From the reasonable perspective, Hobbesian contractarianism violates the constraint of justifiability in two ways. First, it rules out the ‘powerless’, those people who cannot bring benefits to others during social cooperation: children, the severely handicapped.⁵⁶ This exclusion can be attributed to the assumption of mutual unconcern. This assumption does not mean that the contractors are wholly egoistic.

⁵⁶ Lawrence Becker, from a perspective sympathetic to Hobbesian contractarianism, argues that in fact it is mutually beneficial to include most of the disabled persons in the contract. Many disabled people are either already capable of contributing or could be so with accommodation or rehabilitation, and thus it is to the advantage of society to provide accommodations or rehabilitation at some level, requiring then reciprocal contributions from those thus benefitted. For Becker’s suggestion, see Becker (2005).

Moreover, Hobbesian contractarians could further argue that, in general, everyone has a chance to become powerless. People inevitably get sick or becomes handicapped on some occasions. From a rational perspective, in order to avoid being excluded at that time, they should agree upon a social contract that establishes a safety net which protects the powerless. However, these responses are inadequate, because in some cases the difference in power between powerful and powerless may be so large that the powerful people are willing to take the risk after rational calculation. The powerful people may compare the benefit that they might receive after excluding the powerless and the loss that they might suffer if they unfortunately became powerless, and may eventually still find that the benefit can compensate the loss. Also, even though the powerless people are included, they may receive only a small amount of material welfare due to their weak bargaining power, and this is unacceptable from the reasonable perspective. From this we can see that a social contract based on rational bargaining hardly guarantee that each person could be reasonably treated. This is the genuine problem of Hobbesian contractarianism. Whether or not the powerless people are excluded is only the problem on the surface, because even though the powerless people are included, the Hobbesian contract still fails to provide reasonable treatments for each person. I am indebted to Chandran Kukathas for instructive discussion on this point.

Rather, it means that, in most of the bargaining cases, the contractors are only concerned with maximizing their own utilities and are not interested in the utilities of people with whom they bargain. Thus they will cooperate only with people who can help them to improve their utility. Clearly this assumption is not wholly true, for, on some occasions, people are willing to help others at the cost of their own utility. But, Hobbesian contractarians doubt, how frequent are these occasions? In general, the dominant motivation is still the motivation to maximize one's own utility: 'The scope of altruistic motives that are strong enough to normally override self-interest is, for most people, small, that is, confined to concern for family, close friend, close associates...'⁵⁷ Therefore, Hobbesian contractarians believe that the assumption of mutual unconcern is generally true. Another reason for ignoring altruistic motivations is because these motivations are only based on contingent, special affective relationships. Since Hobbesian contractarians aim at justifying *universal* obligation which is independent of any affective ties, these contingent motivations should not be taken into account.⁵⁸ Hobbesian contractarians want to show that, even if we have no affective relationship to one another at all, we still owe to one another certain obligations. Therefore, 'other-interested desires play no role whatsoever in [Hobbesian] justification or explanation of the formation of the state'.⁵⁹

However, if what I said about the generality of reasonableness is true, Hobbesian contractarians surely underestimate how common is the motivation to justify ourselves. In fact, people usually care about whether their behaviour could be justifiable to others, even to those to whom they have no special relationship. This can be shown from the psychological experiments discussed in Chapter 3. The psychological experiments show that many people generally refrain from being the

⁵⁷ Kavka (1986: 65).

⁵⁸ Gauthier (1986: 103). See also Gauthier (1988: 213-214).

⁵⁹ Hampton (1986: 22). Similar claim can also be found in Gauthier (1986: 11).

kind of person who takes justifiability to be insignificant.⁶⁰ Hobbesian contractarians might argue that people in fact do not care so much about justifiability. For example, most of the people always ignore the fact of severe global inequality and keep on spending their money on entertainment rather than spending it to aid the poor. They know that it is unjustifiable, but they still go on doing so. However, according to these psychologists, this is only because people's information is 'distorted' by the distant location of the victims. The motivation of being reasonable is always underestimated only because many people do not receive adequate information in their daily life. If victims appear in an 'up-close-and-personal' manner, say, starving before you, then most people will then be concerned with justifiability. Hence, Hobbesian contractarians' observations are incorrect provided that people are well-informed of the effect of their behaviour. People are also concerned with reasonableness.

From this we can see that the assumption of mutual unconcern is problematic. This assumption is based on a wrong observation that people generally do not care about one another. But the truth is, even if people are unknown to one another, they still care about whether or not their behaviours justifiable. Because of this wrong assumption, the scope of contract in Hobbesian contractarianism is objectionable to people. People naturally care about whether they behave as a moral person who is concerned with justifiability. This respect is unconditional, no matter how much bargaining power others have. Even though the powerless do not have bargaining power and would not bring benefit to others, they still have the capacity for reason-assessing. They are still creatures who are able (or have the potential to be able) to ask for justification from others. Few powerful people would deny that justifiability to the powerless matters and bargaining power is everything. The powerful might still give other reasons to reject the claims of the powerless, but they would rarely

⁶⁰ Greene, Nystrom, Engell, Darely and Cohen (2001)

claim like Thucydides in the Melian dialogue: ‘the strong do what they have the power to do and the weak accept what they have to accept’. Although the powerful do not have any reason to care about the powerless from the rational perspective, they have one from the reasonable perspective.

In light of the importance of reasonableness, we can explain why the scope of a Hobbesian contract is intuitively problematic. In fact, some critics of Hobbesian contractarians have already pointed out that the exclusion of the powerless is a critical flaw of the theory. One such critic is Jean Hampton:

Hobbesian moral theory gives us no reason to respect those with whom we have no need of cooperating, or those whom we are strong enough to dominate, such as old people, or the handicapped, or retarded children whom we do not want to rear, or people from other societies with whom we have no interest in trading...[But] *regardless* of whether or not one can engage in beneficial cooperative interactions with another, our moral intuitions push us to assent to the idea that one owes that person respectful treatment simply in virtue of the fact that he or she is a *person*. It seems to be a feature of our moral life that we regard a human being, whether or not she is instrumentally valuable, as always intrinsically valuable.⁶¹

Moreover, Brian Barry criticizes this approach in a harsher way:

[Hobbesian contractarianism] is by ordinary standards very impoverished. In particular, the ‘congenitally handicapped and defective’ fall outside its protection....I do not believe that when we talk about enabling seriously handicapped people to lead productive lives we mean anything except enabling them to lead lives that are worthwhile to themselves. We are not suggesting that caring them can be made to show profit. That Gauthier thinks we must be

⁶¹ Hampton (1991: 48-49)

claiming this simply shows that he seriously believes other people to occupy his own morally pathological universe.⁶²

These critiques are persuasive, but they do not get to the underlying issues. They simply rely on moral intuitions but do not explain why these moral intuitions are so crucial. If the critiques stay at this level, then Hobbesian contractarians could easily dismiss them by doubting how far these intuitions are reliable: ‘If the reader is tempted to object to [Hobbesian contractarianism], on the ground that his intuitions are violated, then he should ask what weight such an objection can have...’⁶³ However, our discussion of reasonableness provides a concrete ground for these critiques. The fact that reasonableness is an aspect of practical reason can explain why people have strong intuitions to object to political principles that exclude the powerless. For, although those powerless people cannot contribute to the pursuit of people’s interest, they are still able to ask for justification. They are weightless from the rational perspective, but they have weight from a *reasonable* perspective. From the reasonable perspective, there is a strong reason to include the powerless into social cooperation. That is why people have a motivation to reject the Hobbesian contract, even at the cost of their interests. The source of this motivation is not merely a groundless, unreliable intuition, but rather an obligation that is justified by the reasonable aspect of practical reason.

4.4.3 Why Hobbesian contractarianism is to be reasonably rejected? (2): the inadequacy of the Lockean proviso

Apart from the problem with the scope of contract, another reason why Hobbesian contractarianism is reasonably objectionable is that, even if people are included in

⁶² Barry (1995a: 42).

⁶³ Gauthier (1986: 269).

the contract, they can still be unjustifiably treated. Gauthier might disagree with this point by arguing that he also recognizes the importance of reasonableness and the Lockean proviso in his theory represents the constraint of justifiability in reasonableness. Under the regulation of the Lockean proviso, no one in the initial bargaining situation is unjustifiably treated. So I will now discuss whether the Lockean proviso can prevent people from being unjustifiably treated. In Gauthier's contract, predation and parasitism are prohibited in the initial bargaining position, for the Lockean proviso acts as a minimum moral baseline and distinguishes the initial bargaining position from the mutually predatory natural situation. It says that one person should not appropriate some goods if this appropriation 'beters one's situation through interaction that worsens the situation of another'.⁶⁴ Whether a person is better-off or worse-off by the action of another depends on comparing the current situation with the situation in which the other was absent. I am made worse-off by your acquisition if my utility could be higher in your absence. For example, suppose that there is a strong person who does not produce anything but only waits for a weak person to produce and then seizes these goods. This weak person is then made worse off because of the acquisition and hence his rights are infringed by the strong person. Based on this proviso, the rights of people are defined. The bargains, which are based on these rights, are not tainted by predation and parasitism. What goods a person can legitimately own depends on how much they can acquire as long as the Lockean proviso is not violated. This proviso is controversial for it seems that rational utility-maximizers have no reason to accept this proviso if their utility could be maximized through predation and parasitism.⁶⁵ However, whether it is rational to accept the Lockean proviso is not my concern. What I want to discuss is how Gauthier might argue that his contract theory in fact does not ignore justifiability and the Lockean proviso is the evidence.

⁶⁴ Gauthier (1986: 205).

⁶⁵ Representatives of this view are Danielson (1991), Narveson (1991); Goodin (1993); Harman (1988) and Buchanan (1988).

I doubt that the Lockean proviso can perform the function of the constraint of justifiability. For example, something can be clearly unjustified to others but not violate the Lockean proviso. Imagine that there is a rare but highly valuable resource in the world and I am the only one who knows how to exploit this resource. I then appropriate it, and exploit it. Since this resource is highly valuable, I become very rich by selling it. But my appropriation does not violate the Lockean proviso because if I were absent, no one would have had the skill to exploit it and the resource would be left alone. Since no one would be better-off if I were absent, my appropriation does not make anyone worse off. Even if a great gap of wealth between myself and others is created because of my appropriation, this is still permissible by the Lockean proviso. However, this is clearly unjustifiable from the reasonable perspective. Again, the constraint of justifiability requires people to test their behaviour by imagining what a person who is concerned with justifiability would do. Imagine that I was a person who doesn't possess the skills to exploit this resource. I would become relatively poor because the resource is monopolized by people who have the skills to exploit it. Compared with the rich class, I would have a much lower social status and a relatively poor life. Many people find this situation hardly acceptable because of its serious unfairness. The bargaining powers would be radically unequal to the extent that it is difficult to call it a fair distribution. Due to this unfairness, this situation is unjustifiable to many people even though, according to the Lockean proviso, no one was made worse off. From this example we can see how the Lockean proviso fails to rule out behaviours that violate the constraint of justifiability.⁶⁶

There are also behaviours that are clearly justifiable, but the Lockean proviso prohibits them. For example, imagine that there are two fishermen, A and B. Both are working in the same sea area. Given that fisherman A is more talented and

⁶⁶ Similar examples can also be found in Hubin and Lambeth (1991: 119-121) and Lehning (1993: 112-114).

hardworking, they do better and the poor fisherman B does badly. This result frustrates fisherman B. Fisherman B is an envious person and would prefer to be the best fisherman in this sea zone. But A is a contented person who is insensitive to this kind of comparison. Now it appears that the productive activities of fisherman A violate the Lockean proviso. For fisherman B would have a higher utility in the absence of fisherman A. If fisherman A were absent, fisherman B's preference to become the best fisherman in this sea zone could be satisfied and hence his utility could increase. So fisherman B can claim that fisherman A's use of power and ability betters A's position and worsens B's position. However, it seems absurd to say that fisherman A is not justified in using his abilities. Provided that the result of distribution is not radically unequal, there is no problem for A to exercise his talents and become the best fisherman, even though the preference of B will be compromised. B's objection is unjustifiable because even if people were put into the shoes of B, many of them would still think A's use of abilities is allowable. From this example we can see how Lockean proviso sometimes rules out behaviours that are permitted by the constraint of justifiability.

These counterexamples show that behaviours prohibited by the Lockean proviso are significantly different from behaviours prohibited by the constraint of justifiability. One could hardly claim that these two circles are the same. In fact, the Lockean proviso is mainly concerned with the comparison of utility between individuals. No matter what the preferences are and what has occurred, once a person's utility decreases because of the increase of another person's utility, this person is made worse off. However, the constraint of justifiability is concerned with something *different from* this. It prohibits some actions, independent of what the effect of these actions on others' utility would be. These actions are prohibited simply because people generally think that it would be unacceptable if they were treated by these actions. They are unjustifiable to people, but these actions do not necessarily violate the Lockean proviso. For example, the constraint of justifiability is concerned with fairness; it requires the distribution of goods to be fair to all people. The meaning of

fairness cannot be exhausted by comparison between utilities. For sometimes even if people are not made worse off, they can still be unfairly treated in a way that would be unacceptable to anyone who was in their position, as the rare resource example illustrates. As David Copp argues, 'the proviso would permit a rich man to keep his goods, while an unlucky woman starves on his doorstep'.⁶⁷ Hence we can see the Lockean proviso is inadequate from the perspective of reasonableness. The bargaining outcome which is based on the Lockean proviso can still violate the constraint of justifiability.

Therefore we can see that Hobbesian contractarianism is reasonably objectionable in two ways. First, it fails to include the 'powerless' in social cooperation. Secondly, even within the scope of contract, it fails to prohibit behaviours which are generally unjustifiable (and sometimes it prohibits behaviours which are generally justifiable). In these two respects, Hobbesian contractarianism violates the constraint of justifiability. Even though the interests of people are rationally advanced in the mutually beneficial agreement, reasonableness would still motivate people to object to it.

This explains why Hobbesian contractarianism can satisfy the condition of generality only, but fails to satisfy the condition of priority. Since reasonableness is one of the aspects of practical reason, it also has the normative authority to justify or reject obligations, just as rationality does. Ignoring it is a false step of Hobbesian contractarianism. The result is that, even though people will accept Hobbesian contractarianism because of rationality, this is not enough to secure the allegiance of people. The conception of rationality justifies the obligation to follow Hobbesian contractarianism, whereas the conception of reasonableness justifies the obligation to

⁶⁷ Copp (1991: 210).

reject Hobbesian contractarianism. Since both of these obligations have the same strong normative force, this implies that, on the one hand, people have a strong reason to agree with political principles and, on the other hand, people have the same strong reason to disagree with the same political principles. They have no overriding reason for adhering to political principles. Reasonableness will become a source of strong motivation to go against Hobbesian contractarianism.

4.5 Can reasonableness be reduced to a preference?

The objection in the last section explains why people generally have motivations that are contrary to Hobbesian contractarianism, and thus explains why it fails to fulfill the condition of priority. Gauthier might respond that this objection is based on a misunderstanding of the nature of utility. In fact, utility has a formal definition whereby it can include any preferences, including reasonableness, which can be taken as a 'preference to be reasonable'. Thus reasonableness has been taken into account in the utility of an individual. If those people who care about justifiability can understand that their preferences have been satisfied in the optimal outcome already, then they should understand that they are not justified in rejecting the Hobbesian agreement because of justifiability, because their utilities have been maximized subject to the maximization of the individual utilities of others. However, I doubt whether this response works. First, reasonableness has distinctive features which are significantly different from the preferences which are taken into account in the rational calculation of utility. Secondly, if Gauthier insists that following the constraint of justifiability is only a preference in utility function, then the definitions of preference and utility will be broadened to the extent that they go against our ordinary understanding.

4.5.1 The reductionist argument

The reductionist argument is one of the most commonly used arguments when Hobbesian contractarians, like Gauthier, respond to the critiques that they overlook the moral sentiments of human beings.⁶⁸ Gauthier argues that these critiques are based on a narrow understanding of individual utility. Although Gauthier assumes that people basically want to maximize their utilities, it does not mean that people are egoistic and care for nothing but their self-interest. In fact, utility represents only the preferences of a person which are satisfied, and the idea of preference is neutral. That people have a preference means only that they want to attain a particular state of affairs, and this state of affairs is not necessarily a state of affairs in which their self-interest must be satisfied or well-being promoted. Therefore, when Gauthier assumes that human beings generally want to maximize their utilities, it does not mean that human beings care only about pursuing their self-interest. Other non-egoistic, altruistic preferences are also included in individual utility functions. 'If I have a direct interest in your welfare, then...I have reason to promote your welfare'.⁶⁹ Therefore, Gauthier objects to the egoistic interpretation of Hobbesian contractarianism:

[The idea of pursuing individual interests] is unfortunately misleading insofar as it suggests that morality merely constrains egoism; I want to defend morality as a rational constraint on the pursuit of one's aims or objectives, whether or not these objectives have any connection

⁶⁸ Example of this kind of critique are Vallentyne (1991b: 71-75) and Copp (1991: 222-223). Here I have to distinguish my critiques from theirs. These critics, such as Vallentyne, merely disagree with Gauthier that he ignores people's affection for others who have special relationship to them, such as their parents and friends. People are not mutually unconcerned with one another and care not only about their own interests. However, the critique I made in the previous section is different, if not more radical. I argued that people generally want to act in a way which could be justified to others, regardless of whether they have special relationship to one another. This inclination is a general phenomenon of human interactions and does not exist only in the interactions among people who have special affective relationships with one another, thus Gauthier has fewer excuses for ignoring it.

⁶⁹ Gauthier (1986: 7). Similar point can also be seen in Gauthier (1993: 184).

Chapter 4 Hobbesian Contractarianism

with one's interest, or one's personal well-being...the formal aim of the rational individual is the maximum realization of her substantive aims.⁷⁰

Hobbesian contractarians might then acknowledge that being reasonable is a general aspect of the behaviour of people. However, this motivation should not be understood as anything more than a preference which has been taken into account in the rational deliberation. That is to say, people prefer attaining a state of affairs in which they stand in a relationship to one another which is justifiable, and when they attain this state of affairs, their utilities increase. This preference is no different from any other preference that a person possesses, such as the preference to achieve success in business or the preference to have a happy family. Gauthier could argue that this is a better way to understand the conception of reasonableness as a preference which is subject to the conception of rationality, rather than to understand it as independent from rationality.

In this interpretation, reasonableness has been taken into account when people bargain with each other in order to maximize their individual utility. People who prefer attaining a state of affairs in which all people behave justifiably to one another will bargain with those people who do not share this preference. After a long bargaining process, they will reach an optimal agreement that the preferences of these two parties are maximized in a mutually compatible way. This agreement will be the 'principle of minimax relative concession', which means that everyone can minimize the maximum relative concession that they would have to make in order to live with one another. This is a rational solution for everyone because everyone's concession can be minimized. Adopting any alternatives represents that someone has to take unnecessary burden and cannot maximize his individual utility. Those people who have a preference to be reasonable should be satisfied with the optimal

⁷⁰ Gauthier (1991b: 323). Similar point can also be found in Gauthier (1986: 73).

agreement and not seek to attain a more 'reasonable' state of affairs. For breaking the Hobbesian agreement would only make everyone worse-off, including those who prefer attaining a more 'reasonable' state of affairs. Therefore, if those people are rational, they should be satisfied with Hobbesian agreement. By adopting the reductionist argument, Gauthier could avoid the critique that he overlooks the independent moral motivations of human beings by emphasizing the neutrality of their assumptions.

4.5.2 The other-regarding feature of reasonableness

I shall argue that this reductionist argument is unsound because it goes against some fundamental intuitions. First, this argument ignores the distinctiveness of reasonableness. Secondly, this argument relies on a definition of utility that is too broad, if not meaningless. My first objection relies on an observation that reasonableness is clearly different from those preferences which are included in the rational calculation. In general, those preferences are self-regarding, but reasonableness requires people to be other-regarding. They have different natures and the feelings that they arouse are also different. According to Gauthier, although preferences are not only limited to self-interest that are 'interests in the self, that take oneself as object', they must be 'interests of the self, held by oneself as subject'.⁷¹ That is to say, these preferences must be self-regarding. They are necessarily related to myself and are not necessarily related to others. When I have a preference, it must mean that *I* prefer a particular state of affairs, but does not necessarily mean that others prefer this state of affairs as well. However, reasonable behaviour works the other way around: it is by nature, other-regarding. It is necessarily related to *others* and is not necessarily related to *myself*. When I want to be reasonable, what matters

⁷¹ Gauthier (1986: 7).

is that *others* prefer a particular state of affairs and I, due to the constraint of justifiability, should support their preference, even if I do not prefer this state of affairs. People rarely feel that they have to be reasonable only when being reasonable is related to their preferences. Rather, they would think that they have such a commitment, no matter what their preferences are. Here we can see the difference between reasonableness and those preferences which constitute individual utility.

The difference between these preferences is more explicit in the cases where the satisfactions of these preferences are upset. When those self-regarding preferences are upset, a sense of disappointment will appear. One will feel angry and disappointed as long as there is a failure to attain certain states of affairs that one desires. For example, when an athlete takes part in racing and, unfortunately, loses, he will feel disappointed as his dream of becoming the champion breaks down. Anger and disappointment emerge only when self-regarding preferences are upset. However, when one ignores the importance of reasonableness, that is, when one recognizes that one violates the constraint of justifiability, what will appear is not anger and disappointment. Rather, what will appear is guilt and remorse. For example, when people find that they behaved in a way which could not be justifiable to others, they face blame from others and want to do something to compensate for their wrongdoing; what pushes them to compensate should be a sense of guilt and remorse. The feelings of guilt and remorse are distinctively different from the feeling of anger and disappointment. When one feels guilty or disappointed, these feelings are rarely relevant to any preferences for a state of affairs.

Therefore, the sense of guilt should be related to something which has a completely different nature from that of the self-regarding preferences. I believe that the difference between guilt/remorse and disappointment/anger shows that these feelings come from different sources. Disappointment and anger come from self-regarding preferences, which constitute individual utility. Guilt and remorse come from the

other-regarding norm of reasonableness. These two categories of feelings are independent of each other and cannot be reduced to each other. This is why the reductionist argument fails—because it overlooks the feature of other-regardingness. Reasonableness creates a distinctive kind of motivation and one should not be seen as looking for the satisfaction of preferences when one wants to be reasonable.

4.5.3 The cost of broadening the definition of individual utility

Gauthier might avoid this objection by widening the scope of individual utility. He could say that the word ‘preference’ should not be understood literally. A preference does not really mean that a person ‘prefers’ a state of affairs. It means that a person wants to attain this state of affairs, whether or not this state of affair would benefit them. As long as a person *intends* to attain this state of affairs (no matter what the reason is), this intention is taken as a preference and included in the utility function. So the preferences which constitute individual utility are not only self-regarding motivations, but also include other-regarding motivations. Gauthier might remind us that, in the theory of rational choice, the most crucial things are whether a preference is considered and whether a preference is coherent with other preferences.⁷² Whether this preference is self-regarding or other-regarding is much less important. Insofar as an other-regarding preference is considered and coherent, it can be taken into account.

In this most formal interpretation, reasonableness can be reduced to a preference and be included into the utility function. However, the cost of this formal interpretation is also enormous, for this interpretation goes against what we usually conceive about ideas of preference and utility. Normally when we say that our utilities increase, or

⁷² Gauthier (1986: 23-25).

our preferences are satisfied, it usually implies that we become *better-off*. However, we rarely have this feeling when we push ourselves to act in a way which could be justified to others. When we force ourselves to comply with the constraint of justifiability, we usually do not feel that any of our preferences are satisfied. On the contrary, we feel that we become worse-off, not better-off, and our utilities decrease, not increase. We can imagine that, when we save the person whose leg is bleeding in the countryside by driving him to the hospital, we do not expect to become better-off after we help this person. Instead, we clearly understand that we are losing something when we are doing so. For example, the leather upholstery of our car may be damaged; our journey will be interrupted, and so on. However, we still help the injured person. We help them not because we will become better-off after helping the injured, rather; we help simply to avoid unjustifiability, even if this implies that we have to suffer a loss in individual utility.⁷³ We do not think that anything is maximized after we perform this reasonable behaviour. The broad definition of individual utility requires us to think that we are in fact better-off in certain cases, no matter what we really think.

Yet Gauthier might simply bite the bullet and deny that we should take our ordinary understanding of preference and utility for granted. He may cut off the connection between preference and being better-off, and deny that the satisfaction of preferences necessarily implies becoming better-off. Again, a preference means only that a person wants to attain a state of affairs, and this state of affairs is not necessarily a state in which they are better-off. They can intend to attain a state of affairs in which they become worse-off. Nevertheless, even though the definition of preference is loosened to this extent, it is still perhaps odd to reduce reasonableness into a preference, for this will distort the very idea of reasonable behaviour. No matter how

⁷³ It does not mean that cannot make a person better-off. Some people may have a good feeling after helping others. However, people rarely help others *because* they want to derive pleasure from doing so. The pleasure is merely by-product of being reasonable.

Gauthier broadens the definition of preference, he still has to acknowledge that a preference is *consequence-oriented*—‘preference relates states of affair’.⁷⁴ The object of a preference must be a state of affairs. This is the fundamental feature of preference in the theory of rational choice, yet this consequence-oriented feature is absent in reasonable behaviour. Rather, people usually act reasonably *regardless of the consequences*. For example, when people refrain from stealing something, they are not aiming at attaining a state of affairs in which they do not steal anything. This is a strange way to describe their thinking. Rather, their thought should be that ‘no matter what the state of affairs I would be in, I still would not steal’. Considerations about the state of affairs seem to be irrelevant when a person chooses to adhere to the constraint of justifiability.

Gauthier might further respond that reasonableness can still be represented by a ‘preference for a state of affairs in which I follow the constraint of justifiability’. But here the words ‘state of affairs’ are merely redundant, for what ultimately matters is the constraint of justifiability, not the state of affairs that I follow the constraint of justifiability. Again, it seems hard to deny that reasonableness is significantly different from the preferences that are included in the rational calculation: the latter takes states of affairs as objects, while the former does not.

Therefore, the reductionist argument fails to rescue Hobbesian contractarianism because it cannot capture the distinctive features of reasonableness. Reasonableness cannot be reduced to a preference in the rational calculation because of its other-regarding feature. Hobbesian contractarians could further loosen the definition of preference in order to include other-regarding preferences, but the cost of this will be to go against our ordinary usage of preference and utility. The failure of the

⁷⁴ Gauthier (1986: 22).

reductionist argument shows that reasonableness has an independent status in our practical reasoning. It cannot be reduced to any preferences, thus it is ignored by Hobbesian contractarians who are concerned only with justifying a mutually beneficial agreement. But reasonableness is also fundamental since it is also a conception which governs practical reasoning. Hence, the existence of reasonableness as an aspect of practical reason causes problems for Hobbesian contractarianism, for even if people benefit by a Hobbesian agreement, they might still refuse to support this agreement because of the unjustifiability of this agreement.

4.6 Conclusion

Hobbesian contractarianism is undoubtedly one of the most commonly adopted approaches of contractarianism in contemporary political philosophy. This can be attributed to two virtues of the approach. First, this approach can capture some genuine and fundamental features of our daily decision-making and hence it can fulfill the condition of generality. People usually intend to pursue their interests rationally. If certain obligations which are imposed on a person are in their rational interest, then even if they have not consented to them, we can still predict that they would probably consent if they were asked. That is why readers rarely think that the Hobbesian hypothetical contract is a pointless story. Secondly, this contractarian approach can make use of game theory to develop clear conclusions. Hobbesian contractarians exploits the strength of the theory of rational choice to facilitate a deep and systematic analysis of interaction among rational individuals.⁷⁵ As Jody Kraus notes, ‘their detailed and sustained analysis of interaction in Hobbesian hypothetical

⁷⁵ As Brian Barry points out, the development in rational choice theory, for example, John Nash’s research on bargaining problems helps us a lot in predicting determinate solutions to the bargaining problem between rational agents. Until the rational choice theory was well-developed, bargaining problems were usually regarded as having no determinate solution. See Barry (1989: 12).

settings is indeed impressive'.⁷⁶ This level of clarity is rarely seen in other contractarian theories. For example, Kantian contractarianism is notorious for its ambiguity because it fails to derive a definite conclusion which would be achieved in its initial choice situation.⁷⁷ Compared with Kantian contractarianism, either the conclusion or the inference process of Hobbesian contractarianism is highly clear, since it borrows force from the rigorous mathematical models. These two features of Hobbesian contractarianism can explain why so many contractarians are impressed by this approach.

Nevertheless, these merits are achieved at the cost of simplifying the complex structure of practical reason. Due to this simplification, Hobbesian contractarianism fails to fulfill the condition of priority. Since Hobbesian contractarians want to exploit the strength of game theory, they have to adopt the assumption that human beings are by nature rational utility-maximizers and regard the other features of human beings as irrelevant or reducible. This assumption is not wholly wrong, for rationality is generally acceptable to actual people. However, this assumption is not wholly correct, for people are not *only* rational. Apart from rationality, reasonableness is also another aspect of practical reason. Since the Hobbesian agreement is unjustified from a reasonable perspective, reasonableness becomes a source of strong motivation to push people to be 'unjust', and Hobbesian contractarianism fails to explain how their hypothetical contract can account for this motivation. Therefore, the normativity of Hobbesian contractarianism is unsatisfactory, because people will not take Hobbesian agreement as having the

⁷⁶ Kraus (1993: 316).

⁷⁷ Aaron James mentions this critique of Scanlon's theory. See James (2004: 369-377). Although Barry's theory, which is also a Kantian contractarian model, offers more specific political principles (such as the principle of equal religious freedom), it is vulnerable to another problem. The inference process of these principles is too rough and it is always unclear why reasonable people would choose these principles but not the others. This flaw can be attributed to Barry's vague definition of reasonableness and I will further explain it in Section 5.1.1.

highest priority. As Albert Weale observes, the strength of Hobbesian contractarianism is its instrumental conception of rationality, but at the same time, it is the flaw of this theory as well, and this flaw indicates that ‘the contractarian enterprise’ calls for a more comprehensive view of practical reason.⁷⁸

In fact, most of the critics of Hobbesian contractarianism focus too much on the incoherence within the theory, and place too little importance on its flawed psychological assumptions.⁷⁹ When theorists discuss the problem of Hobbesian contractarianism, they concentrate on the irrationality of constrained maximization. Gauthier proposes that a rational utility-maximizer should be a constrained maximizer, which complies with the contract unconditionally. Most of the critics are unsatisfied with this point and argue that a rational utility-maximizer should instead obey the mutually beneficial agreement *conditionally*. They have no reason not to break the rules if this will not be discovered by others. Hence, the most rational choice is to be a careful rule-breaker, rather than be a rule-follower. Surely this is a serious problem of Hobbesian contractarianism, yet I fear these critics accept the assumptions of Hobbesian contractarianism too easily. At least they have not examined these assumptions carefully enough. Only some argue that Hobbesian contractarianism goes against our ordinary moral intuition, but they do not discuss it in depth but merely take it as showing that Hobbesian contractarianism is counter-intuitive.⁸⁰ Yet this kind of critique does not have much power because one of the fundamental beliefs of Hobbesian contractarians is that fragmented moral intuition is contingent and unreliable, so morality needs to be based on a general foundation.⁸¹ However, in

⁷⁸ Weale (1993).

⁷⁹ See Kraus (1993), Hampton (1991), Copp (1991), Sayre-McCord (1991), Kraus and Coleman (1991), Nida-Rumelin (1993), Sugden (1993), Finkelstein (2001) and Harman (1988). For a response of the incoherent critique, see McClennen (2001). For a Gauthier’s attempt to answer this problem, see Gauthier (1982) and Gauthier (1984).

⁸⁰ See Hampton (1991) and Barry (1995a).

⁸¹ Gauthier (1986: 269).

this chapter, I showed the real strength of this critique. The intuition is not, as some Hobbesian contractarians think, ‘tentative and controversial’⁸² but rather has its roots within an aspect of practical reason, which pushes people to follow the constraint of justifiability. This aspect of practical reason is not any less important than rationality. In the next chapter, we will see another approach of contractarianism which is based on that aspect of practical reason and develops another completely different model of contractarianism.

Finally, I would like to end this chapter with a story. When I was writing this chapter, I went to a conference which was for the 25th Anniversary of the publication of *Morals by Agreement*. At the end, the author, David Gauthier, gave the closing speech. Since he has retired and this conference was likely to be his last academic activity, the speech can be seen as his final conclusion on the project that he has pursued his whole life. In the speech, he proposed that it is misleading to assume that hypothetical contractors are rational utility-maximizers. Rather, contractors should be assumed to be rational utility-*optimizers*, which mean that they are concerned with maximizing their utilities *on condition that this would be acceptable to others*. Maximization is only one of the considerations contractors care about. Contractors also care about whether their acts could be justifiable to others, which is exactly something that *reasonableness* emphasizes. This change is ironic, for it seems that the most representative Hobbesian contractarian now also recognizes the inadequacy of the instrumental conception of rationality and the importance of justifiability.⁸³ Rational interest is crucial, yet it is not the only important issue.

⁸² Gauthier (1986: 270).

⁸³ This can also be further supported by the observation of Robert Brandom: ‘in the years since *Morals by Agreement* appeared, Gauthier has in many ways moved in a Kantian direction’. See Brandom (2001: 33).

Chapter 5 Kantian Contractarianism: Scanlon and Barry

5.1 Introduction

In her famous article ‘Moral Beliefs’ Philippa Foot mentions a debate between Socrates and Thrasymachus in the *Republic*.¹ Thrasymachus raised a provocative question: ‘Why should I be just?’ He argued that being *just* is in fact not something which should be taken for granted. He began with a premise that, in certain circumstances, injustice was more profitable than justice. This premise ended with a conclusion that a person with strength to get away with injustice has reason to follow this as the best way of life. This argument is provocative because it is an argument which justifies injustice and, according to Foot, every moral and political philosopher should ask how they can respond to this argument. Interestingly, Hobbesian and Kantian contractarians give opposite answers to this question. In order to refute this argument, Hobbesian contractarians choose to reject the premise. They deny that injustice is more profitable than justice; therefore rational people should behave justly for the sake of maximizing their individual utility.

In contrast with Hobbesian contractarianism, Kantian contractarianism, which is the focus of this chapter, accepts the premise of Thrasymachus, but rejects his conclusion. According to Kantian contractarian, the conception of reasonableness, but not the conception of rationality, is the dominant aspect in practical reason. That is to say, even if violating political principles could serve a person’s utility better than adhering to them, a just or moral person would still comply with them because of reasonableness. Based on this different assumption, Kantian contractarians develop their contract model in a way different from that of Hobbesian contractarians. As Scanlon says,

¹ Foot (2002: 125-127).

Chapter 5 Kantian Contractarianism

‘Contract’ and its cognates seem to many people to suggest a process of self-interested bargaining that is foreign to my account. What distinguishes my view from other accounts involving ideas of agreement is its conception of the motivational basis of this agreement. The parties whose agreement is in question are assumed not merely to be seeking some kind of advantage but also to be moved by the aim of finding principles that others, similarly motivated, could not reasonably reject.²

This strand of contractarianism, which is represented by T. M. Scanlon and Brian Barry, is called Kantian contractarianism because these thinkers are highly indebted to the ideas of Kant (and sometimes Rousseau). As Scanlon acknowledges:

The idea of a shared willingness to modify our private demands in order to find a basis of justification that others also have reason to accept is a central element in the social contract tradition going back to Rousseau. One of the main reasons for calling my view ‘contractualist’ is to emphasize its connection with this tradition.³

The similarity between Kant and Kantian contractarianism can be seen in their shared emphasis on the idea of justifiability. The constraint of justifiability, which is the absolute criterion in the conception of reasonableness, is very similar to the categorical imperative in Kant’s ethical theory. To Kant, when we act in a certain way, we commit ourselves to certain ‘maxims’, which means that our will is placed to under a subjective volitional principle.⁴ We then have to ask ourselves whether we can will this maxim to be a universal law, in order to see whether it could be publicly justifiable. That is to say, we have to ask whether or not a world in which everyone could act on, is possible. If such a world is possible, then my maxim is publicly justified and we would justifiably act on it. This kind of maxim is called the categorical imperative. The categorical imperative test undoubtedly affects Scanlon. He also admits that his idea of reasonable rejectability ‘does have an obvious

² Scanlon (1998: 5).

³ Scanlon (1998: 5).

⁴ Kant (1998: 31).

similarity to Kant's Categorical Imperative'.⁵ To Scanlon, asking whether an act would be allowed by principles that no one could reasonably reject, is highly similar to asking whether an act could pass through the categorical imperative test.⁶ Based on this shared concern with justifiability, Scanlon aligns himself with Kant and believes that he inherits from Kant a distinctive model of social contract—a hypothetical contract that is governed, not by mutual advantage and self-interest, but by mutual respect among free and equal persons.

Like Hobbesian contractarianism, the brand of Kantian contractarianism is also disputable. Some critics argue that it overlooks Kant's 'two worlds' distinction.⁷ Kant is famous for arguing that there are two realms: the realm of nature and the realm of freedom. While people are not free in the former realm for they are subject to the laws of causality, these people, as rational beings, are free in the latter realm for they can stand in a noumenal standpoint and determine the actions of their phenomenal selves. The emphasis on transcendental freedom is one of the most distinctive features of Kant's ethical theory, yet this idea of freedom is generally avoided by Kantian contractarians. They usually try to discard the 'transcendental' characteristics of Kant's theory.⁸ Hence, some people might argue that 'Kantian contractarianism' is the wrong label since there are too few ideas shared among Kant and Kantian contractarians. Again, I will not go into this question in this chapter for my interest is in the normative force of Kantian contractarianism, rather than the relationship between Kant and Kantian contractarianism. Also, despite its controversy, this label has already been widely used in the relevant literatures.⁹ I will focus in this chapter on discussing theories of Kantian contractarians, such as Scanlon and Barry—regardless of what their arguable connection with Kant may be.

⁵ Scanlon (1998: 5).

⁶ Scanlon (2008: 98, 117).

⁷ Flikschuh (2000: 50-51).

⁸ Apart from this, some Kantian contractarians, such as Scanlon, also disagree that acting in certain way represents committing oneself to certain maxim, a kind of law-making activity. It can simply represent choosing an end. See Scanlon (2008: 97-98).

⁹ See Sayre-McCord (2000), Hampton (1993), Kymlicka (1993) and Southwood (2010).

5.1.1 A brief comparison between Scanlon and Barry

At first glance, if the object of our discussion is the justification of political principles, it seems that Barry's contract theory should be the major subject of our analysis, for Barry defines his project as 'a theory about the kinds of social arrangement that can be defended',¹⁰ while Scanlon is more concerned about 'a narrow domain of morality having to do with our duties to other people'.¹¹ Simply speaking, Barry is more 'political' while Scanlon is more 'moral'.

However, it does not mean that Scanlon should be excluded in the discussion of Kantian contractarianism. On the contrary, Scanlon's theory is a valuable resource which can help us to understand the conception of reasonableness and why this conception should have the highest priority. First, compared with Barry, Scanlon provides a more detailed account of the conception of reasonableness. Barry's contractarian theory can be conceived as an elaboration of the earlier Scanlon's approach. As Barry himself admits, 'I cannot therefore see any alternative to returning to the idea of reasonable agreement itself and trying to elaborate it'.¹² When Barry's *Justice as Impartiality* was published in 1995, the conception of reasonableness came mainly from the account provided by Scanlon's famous 1982 article 'Contractualism and Utilitarianism'. However, this account of the conception of reasonableness is unclear and inadequate since, to Scanlon, his article is 'only an outline' and 'much more needs to be said to clarify its central notions'.¹³ Scanlon does not give very much content to the conception of reasonableness and the word 'reasonable' is mostly used in a literal way. It was not until 1998, when *What We Owe to Each Other* was published, that the idea of reasonableness was explained in a more precise and philosophical way. Hence, Barry's contractarian theory suffers from the problem of vagueness because its account of the conception of

¹⁰ Barry (1989: 3).

¹¹ Scanlon (1998: 6).

¹² Barry (1995a: 10).

¹³ Scanlon (1982: 150).

reasonableness is merely a ‘primitive’ account that comes from Scanlon’s earlier writings.

Secondly, although the word ‘justice’ and ‘social institution’ rarely appear in *What We Owe to Each Other*, this does not mean that the question of what social institutions are justified to people falls outside Scanlon’s concern. In a paragraph devoted to *taxonomizing* these matters, Scanlon says that, while he is not discussing morality in the broadest sense, but only a part of it; that part, what we owe to each other, is ‘broader than justice, which has to do particularly with social institution’.¹⁴ This implies that justice and social institutions do fall within his remit. Also, Scanlon believes that his contract theory can provide concrete ground for various social and political rights.

Claims about rights, like other claims about what we owe to each other, are claims about the constraints on individual action, and on social institutions, that people can reasonably insist on. In order to decide what rights people have, we need to consider both the costs of being constrained in certain ways and what things would be like in the absence of such constraints, and we need to ask what objections people could reasonably raise on either of these grounds.¹⁵

This reveals the ‘political’ dimension of Scanlon’s contract theory. Like moral principles, political principles can justify coercing people, but this depends on whether these principles represent agreements that no one could reasonably reject.¹⁶

Therefore, in this chapter, I will take Scanlon’s contract theory to be the main representative of Kantian contractarianism, and treat Barry’s contract theory as a supplementary example. This chapter falls into three parts. The first part presents a

¹⁴ Scanlon (1998: 6-7, 163).

¹⁵ Scanlon (2003a: 4).

¹⁶ One of the examples is that a state can coerce its citizens to uphold contracts, because individuals are morally required to keep their promises. So the power of the state can be legitimately used to force them to do these things. See Scanlon (2001: 249).

brief sketch of Kantian contractarianism and explains why it satisfies the condition of generality. In the second part I argue that, since Kantian contractarianism underestimates the importance of rationality, it fails to fulfil the condition of priority. Unlike Hobbesian contractarianism, which tends to reduce reasonableness to a preference, Kantian contractarianism acknowledges that rationality has an independent status, but it is always subject to the authority of reasonableness. However, rationality is actually a conception of practical reason that has the same importance as reasonableness, so it is unrealistic to assume that the latter can always override the former. I then examine different defences of the priority of reasonableness which are provided by Kantian contractarians, three from Scanlon and one from Barry. I finally conclude that, insofar as all of these defences fail, Kantian contractarianism cannot avoid failing to fulfil the condition of priority.

5.2 Kantian contractarianism: an elaboration

In the last chapter, we saw how Hobbesian contractarians develop a contractarian theory from the assumption of rationality. Unlike Hobbesian contractarianism, Kantian contractarianism relies on an assumption that the most primary aspect of practical reason is reasonableness, not rationality. This section will discuss what kind of contractarian theory can be developed on the basis of this assumption.

5.2.1 Characteristics of hypothetical contractors

While rationality is a formal process that simply requires people to satisfy their self-regarding preferences effectively, reasonableness is a formal process that presupposes an *other-regarding* aim of ‘living with others who are also reasonable’.¹⁷ Kantian contractarians believe that reasonableness constitutes the structure of practical reason of actual people and defines what actual people ought to

¹⁷ Scanlon (1998: 154).

do. Based on this conception of practical reason, justified political principles should be principles which could be reasonably justified. These principles can be specified by a hypothetical agreement that would be made by contractors who were wholly reasonable.¹⁸ In general, the characteristics of hypothetical contractors can be described under three aspects: their motivation, their qualification, and their knowledge.

The motivation of contractor

The account of the motivation of contractors is based on a Kantian understanding of human nature. By contrast with Gauthier, who assumes that hypothetical contractors are merely ‘animals with preferences’, Scanlon assumes that hypothetical contractors have the capacity for ‘legislating’, that is, they could control their desires within a limit that could be justifiable to others.¹⁹ According to Scanlon, a human being is ‘a reasoning creature—one that has the capacity to recognize, assess, and be moved by reasons’.²⁰ He or she can give reasons to justify acts and understand reasons given by others. Since hypothetical contractors are the representatives of actual people, they also possess this distinctive capacity. Because of this capacity, contractors respect each other as creatures that are capable of asking for justification, ‘a nonderivative source of reason’.²¹ Therefore, when contractors participate in a scheme of social cooperation, they would not simply take this scheme as a means to pursue their own

¹⁸ Scanlon might deny that the idea of agreement plays any fundamental role in his argument. For his theory is concerned with what no one could reasonably reject, but not what people hypothetically would or would not agree under certain conditions. See Scanlon (2004: 133-134). Nevertheless, I do not see a big difference, since something that no one could reasonably reject should also represent something that everyone should reasonably agree on, and this could be known by studying what reasonable contractors would agree under certain conditions. Hence, the idea of agreement still plays a fundamental role in Scanlon’s theory. In fact, Scanlon merely wants to emphasize that his hypothetical agreement is not only a thought experiment, a descriptive claim, but rather a normative claim. At this point Scanlon does not have a big difference from other contractarians such as Gauthier and Rawls, who also believe that their hypothetical contracts are not merely descriptive claims but also have normative power.

¹⁹ Scanlon (2008: 97).

²⁰ Scanlon (1998: 23).

²¹ Scanlon (2008: 92).

interests. Rather, contractors share a common aim of ‘living with others who are also reasonable’ and seek a scheme that can fulfil this aim.²² This aim is the origin of motivations that move reasonable contractors to seek principles to which everyone would agree.

When we say, in the course of an attempt to reach some collective decision, that a person is being unreasonable, what we often mean is that he or she is refusing to take other people’s interests into account. What we are claiming is that there is a reason to take these interests into account *given* the supposed aim of reaching agreement or finding a course of action that everyone will happy with.²³

Because of this shared aim, contractors would constrain the pursuit of their interests within the ‘constraint of justifiability,’ that is, the pursuit of interests should be given up if these interests could not be justifiable to other reasonable people. Violating this constraint is the same as harming the relationship of mutual respect, which is ‘the fundamental contractualist moral relation’.²⁴ Hence, compared with the motivation of pursuing interests, contractors have strong ‘standing intentions to regulate their behaviours in certain ways’.²⁵

The qualification of contractors

The scope of the Kantian contract is much larger than that of the Hobbesian contract, for in the Hobbesian contract, moral standing is determined by bargaining power, but in the Kantian contract, moral standing is determined by the capacity for assessing reason. ‘The scope of [Kantian contractarianism] will include those beings to whom we have good reason to want our actions to be justifiable’.²⁶ That is to say, those

²² Scanlon (1998: 192). See also Scanlon (1993: 182).

²³ Scanlon (1998: 33).

²⁴ Darwall (2009: 302); similar point can also be found in Scanlon (2008: 6, 141).

²⁵ Scanlon (2008: 140).

²⁶ Scanlon (1998:179).

severely handicapped can also be included in the scope of contract provided that they are conscious of asking for justification. This boundary seems to exclude children and future generations, but actually they are still included since they ‘will’ have the capacity for reason-assessing.²⁷ However, animals are not included, for human beings are the only beings who possess the distinctive capacity for reason-assessing and this capacity distinguishes human beings from other animals. ‘This reflective capacity sets us apart from creatures who, although they can act purposefully...cannot raise or answer question whether a given propose provides adequate reason for action’.²⁸ Hence, a relationship formed between a human being and an animal is different from a relationship formed between two human beings. In a relationship between a human being and an animal, the human being need not worry about whether his or her action could be justifiable to that animal.²⁹

²⁷ Scanlon (1998: 185). Yet this point is still objected by some theorists as inadequate, see Hooker (2000: 66-70).

²⁸ Scanlon (1998: 23).

²⁹ But it does not mean that human being can do whatever they want to animals, because they have to care about whether their action could be justifiable to those human beings who care about those animals. See Scanlon (1998: 219-220).

The knowledge of contractors

Similar to Gauthier, Scanlon is dissatisfied with Rawls' assumption of the veil of ignorance. Although Scanlon agrees that particular information is relevant when contractors decide what contract should be made, he argues that the veil of ignorance is unnecessary.³⁰ After Rawls shows that principles of justice would be chosen by rational individuals behind the veil of ignorance, he then has to further explain why actual people have to accept the decision which is made in the absence of particular information. This makes the so-called veil redundant, because the reasons why actual people would accept the decisions taken behind the veil would simply be the reasons for directly choosing the principles of justice. Hence, Scanlon suggests that the veil of ignorance 'is not necessary...since the requirement of justifiability (or of nonrejectability) already requires one to take these others into account'.³¹ All hypothetical contractors are familiar with their particular information, so the agreement is an agreement which would be made by people who were fully informed. Scanlon believes that this kind of agreement can capture the strength of social contract theory, while avoiding the problem of irrelevance cause by the veil of ignorance.

5.2.2 The state of nature: the reason-exchanging conversation

These three characteristics constitute the hypothetical contractors in Scanlon's theory—well-informed reasonable persons who respect one another as reason-assessing beings and follow the constraint of justifiability. Scanlon believes that the choices of these contractors represent the choices that actual people would make when they were fully reasonable and not affected any irrelevant factors. Then Scanlon further asks: what political principles will be agreed among these

³⁰ Scanlon (1982: 124-126).

³¹ Scanlon (1998: 207).

hypothetical contractors? For the sake of reaching a consensus, these people go into conversations with one another.³²

In their conversations, hypothetical contractors will propose different principles. Moreover, they do not simply suggest principles. They need to give generic reasons to justify their suggestions, for every hypothetical contractor is reason-assessing creature. Generic reasons are 'reasons that we can see that people have in virtue of their situation, characterized by general terms, and such things as their aims and capacities and the conditions in which they are placed'.³³ Generic reasons refer to reasons which are generally acceptable from various points of view.³⁴ They are based on 'commonly available information about what people have reason to want,' rather than being based on 'the particular aims, preferences, and other characteristics of specific individuals'.³⁵ Since the latter reasons are only acceptable to particular people, they cannot be used as general justifications for political principles. Only generic reasons can be used as currency of reason-exchanging. The whole conversation is an exchange of generic reasons among people. A hypothetical contractor has to weigh up the reasons to accept or reject principles against the reasons of others. If there are some other contractors who have stronger reasons to reject a proposed principle, then it will not be reasonable for them to insist on this principle. Contractors who have a weaker reason should give way to contractors who have a stronger reason.

However, how can we measure the strength of a generic reason? Scanlon suggests that 'we need to first form an idea of burdens that would be imposed on some people [if the principle in question were rejected/accepted]'.³⁶ For example, suppose that a

³² Kumar (1999: 277-278) and O'Neill (2004: 29).

³³ Scanlon (1998: 204). See also Scanlon (1975: 76).

³⁴ Scanlon (2003b: 182).

³⁵ Scanlon (1998: 204).

³⁶ Scanlon (1998: 195). See also Scanlon (1993: 182). 'Burden' can be roughly defined as the loss of one's well-being. Although Scanlon argues that, apart from well-being, other considerations can be counted as grounds for reasonable rejection, still, 'components of well-being figure prominently as

principle is supported by a generic reason but, at the same time, is rejected by a generic reason. If this principle became the fundamental political principle, then person A would have to suffer a burden. On the other hand, if that principle were rejected, then person B would have to suffer a burden. Suppose that, compared with the burden which B would have to suffer if the principle were rejected, the burden which A would have to suffer if the principle were accepted were merely insignificant. Then the principle should be accepted, for the generic reason which supports the principle is stronger than the generic reason which rejects the principle. Therefore, the strength of a generic reason depends upon the burden which would be imposed on someone if this generic reason were rejected.³⁷ The larger the burden is, the stronger the generic reason is. Through this evaluative standard, contractors can compare different principles and finally arrive at principles that are supported by the strongest generic reason.

5.2.3 The hypothetical contract: principles which no one could reasonably reject

According to Scanlon, the content of the hypothetical contract should be principles which were supported by the strongest generic reason, that is, the agreement which 'is least unacceptable to the person to whom it is most unacceptable'.³⁸ Finally,

grounds for reasonable rejection'. For the discussion of the ground for reasonable rejection, see Scanlon (1998: 213-216). For an interpretation of Scanlon's idea of burden as loss of well-being, see Ashford (2003: 280).

³⁷ A similar thought experiment is discussed when Scanlon talks about how we know a right is justified. In fact these thought experiments are not wholly imaginary since they need empirical studies to support their prediction. 'Neglecting this empirical element leads rights to degenerate into implausible rigid formulae'. See Scanlon (1978: 35). Actually this is also another piece of evidence that Scanlon's contract theory can be applied in discussions of political issues.

³⁸ Kumar (1999: 294).

people would come up with a principle which was supported by the strongest generic reason, which would be a 'principle which no one would reasonably reject'.³⁹

This seems to be quite similar to the 'principle of minimax relative concession' proposed by Gauthier, which is also the least unacceptable outcome after contractors bargain in Hobbesian's initial situation. 'Since one's ground of complaint can be measured by one's relative concession, minimizing maximum relative concession minimizes the grounds of complaint'.⁴⁰ Nevertheless, the idea of burden in Gauthier's contract is more restrictive than that in Scanlon's. To Gauthier, only the loss in utility is a burden. Hence, the principle of minimax relative concession is a principle whereby the loss of the worst-off in utility is minimized. But Scanlon's definition of burden is more complex. It is 'an ethically significant, objective notion' which does not only include subjective preferences.⁴¹ Although he acknowledges that 'gains or losses in well-being (relief from suffering, for example) are clearly the most relevant factors in determining whether a principle could or could not be reasonably rejected',⁴² well-being (or utility) is not the unique concern in defining a burden.⁴³ Apart from well-being, many other considerations could also constitute a burden on a person. For example, people could suffer from a burden if they were treated unfairly,⁴⁴ or if their freedom or autonomy were undermined.⁴⁵ Scanlon himself denies that it is possible to give 'a clear specification of the possible grounds for reasonably rejecting a principle (whether this in terms of a conception of well-being or in some other way)'.⁴⁶ Whether certain considerations are or are not relevant to the

³⁹ Kantian contractarians assume that whether a principle could be reasonably rejected is obvious and people should have no disagreement on this issue. See Barry (2000: 191) and Scanlon (1982: 121).

⁴⁰ Vallentyne (1991c: 8-9).

⁴¹ Scanlon (1978: 27) and (1975: 72).

⁴² Scanlon (1998: 215).

⁴³ Scanlon (1998: 213-218). Also, Scanlon's definition of well-being is also broader than Gauthier's definition of utility, which only consists of subjective preferences. See also Scanlon (1975), Scanlon (1978) and Scanlon (1993).

⁴⁴ Scanlon (1998: 212-213, 216).

⁴⁵ Scanlon (1998: 203).

⁴⁶ Scanlon (1998: 217).

reasonable rejectability of a principle depends on our 'normative judgment'.⁴⁷ Although the whole account of burden is a little vague, at least it is clear that well-being is not the only concern when contractors reasonably reject certain principles.

We can take religious freedom as an example, which is discussed by Barry when he talks about what political principles would not be reasonably rejected.⁴⁸ In certain conditions, protecting freedom of religious worship is not a mutually advantageous practice. When one religious group is very strong and another is very weak, the cost for the former to force the latter to follow their religion would be negligible. However, the weaker religious group could suggest other generic reasons apart from utility, such as unfairness. Putting aside the difference in bargaining power first, the weaker religious group could ask a question 'How would you like it if someone did that to you?' While the strong religious group also takes its religion seriously, it should also understand the importance of religious worship. It does not prefer a state in which religious worship would be prohibited. Hence, this means that the stronger religious group is depriving the weaker of something of which they also understand the importance. The weaker could then claim that the rejection of religious freedom is unfair because the stronger 'wouldn't want others to do it to him'.⁴⁹ This appeal to fairness is a strong generic reason that can override the loss in utility and can provide powerful support for the proposal of equal religious freedom. From this example we can see that, although both of them talk about burdens, because of Kantian contractarians' more complex definition of burden, they will reject some principles that Hobbesian contractarians would not.

⁴⁷ Scanlon (1998: 218).

⁴⁸ Barry (1995a: 83-85). Scanlon also emphasizes the importance of religious freedom, even though he does not argue it through a Kantian contractarian framework. See Scanlon (1996). Apart from religious freedom, Scanlon also says that his contract theory can justify certain 'familiar civil rights,' such as freedom of speech, rights of privacy and the right to life,' see Scanlon (1972: 24-25) and Scanlon (1978: 28).

⁴⁹ Barry (1995a: 83).

Therefore, from this hypothetical contract, we can see what (relatively clear) political principles can be derived from a vague assumption of reasonableness. Scanlon begins with an assumption of reasonableness. He then defines the hypothetical contractors for the sake of making this feature of practical reason salient. Through a reason-exchanging conversation, these contractors agree upon principles which no one could reasonably reject, that is, principles which would impose the least burden on the worst-off people.

5.3 Kantian contractarianism and the condition of generality

Scanlon presents a hypothetical contract which justifies certain political principles through the conception of reasonableness. However, is this conception generally acceptable? Why should actual people care about the decision of these hypothetical contractors? If contractarians fail to show that the practical reasoning of hypothetical contractors is generally acceptable, then the political principles are merely coercive. The purpose of this section is to show that, similar to Hobbesian contractarianism, Kantian contractarianism can also satisfy the condition of generality; and this can be shown by considering some common psychological emotions.

5.3.1 The conception of reasonableness and the sense of guilt

Some critics argue that Scanlon does not clearly show the relationship between his hypothetical contract and the real world. Therefore he ‘must explain why [reasonableness] has a deep motivational place in our lives’.⁵⁰ However, Scanlon does not think it is a problem, for he thinks that the aim of ‘living with other reasonable people,’ which is the aim presupposed in the conception of reasonableness, is widely shared to the extent that it gives all of us a direct reason for action. This aim explains why people always have a motivation to be reasonable.

⁵⁰ Watson (1998: 261).

According to the version of contractualism that I am advancing here our thinking about right and wrong is structured by a different kind of motivation, namely the aim of finding principles that others, insofar as they too have this aim, could not reasonably reject. This gives us a direct reason to be concerned with other people's point of view: not because we might, for all we know, actually *be* them, or because we might occupy their position in other possible worlds, but in order to find principles that they, as well as we, have reason to accept.⁵¹

Furthermore, Scanlon offers a phenomenological argument and shows that the conception of reasonableness is widely adopted by people.⁵² He argues that actual people generally take reasonableness as the primary aspect of practical reason which has an absolute priority over other considerations, such as rational interests. The priority of reasonableness can be seen from how people value the relationship of mutual recognition, that is, an interpersonal relationship by which human beings recognize the distinctive features of one another and respect each other as reason-assessing creatures. According to Scanlon, this relationship 'is seen as playing a fundamental role in our moral thinking'.⁵³ People generally think that this relationship is appealing in itself:

The contractualist ideal of acting in accord with principles that others (similarly motivated) could not reasonably reject is meant to characterize the relation with others the value and appeal of which underlies our reasons to do what morality requires...Standing in this relationship to others is appealing in itself—worth seeking for its own sake. A moral person will refrain from lying to others, cheating, harming, or exploiting them, 'because these things are wrong'. But for such a person these requirements are not just formal imperatives; they are aspects of the positive value of a way of living with others.⁵⁴

This relationship is appealing because it represents an ideal moral community where the value of human beings is respected. In Scanlon's words, it is a 'Kingdom of ends

⁵¹ Scanlon (1998: 191).

⁵² Scanlon (1998: 157).

⁵³ Scanlon (2008: 98).

⁵⁴ Scanlon (1998: 162).

on the cheap'.⁵⁵ In this ideal moral community, the distinctive capacities of human beings are recognized and therefore all are respected as *ends-in-themselves*. Scanlon believes that the attractiveness of this moral ideal is apparent and its attractiveness gives us a strong reason to think that the relationship of mutual recognition is exceedingly important.

I *do* find the idea of moral community very appealing. So I try to develop a notion of morality which simply takes as basic that notion of moral community. When you ask 'why be moral?', I think we can just describe the appeal of that kind of community and the dis appeal of its alternative.⁵⁶

The fact that actual people generally value the relationship of mutual recognition can also be proved by the sense of guilt. Scanlon argues that, if actual people fail to stand in a relationship of mutual recognition with one another, actual people will suffer from a strong sense of guilt. 'Feeling guilty for something one has done is plausibly understood as feeling that it has impaired one's relationship with certain people. In my experience there is nothing weak or mild about such a feeling'.⁵⁷ This sense of guilt is so great that 'most people are willing to go to considerable lengths, involving quite heavy sacrifices, in order to avoid admitting the unjustifiability of their actions'.⁵⁸ The power of guilt indirectly reveals the positive value of the relationship of mutual recognition; it 'testifies...to the value people set on the belief that their lives and institutions are justifiable to others'.⁵⁹

According to the account I am offering, the pain of guilt involves, at base, a feeling of estrangement, of having violated the requirements of a valuable relation with others. So

⁵⁵ Scanlon (2009: 189).

⁵⁶ Scanlon (2009: 190). See also Scanlon (2008: 98).

⁵⁷ Scanlon (2008: 157).

⁵⁸ Scanlon (1982: 117).

⁵⁹ Scanlon (1998:163).

understood, this familiar negative aspect of morality corresponds to a positive ‘pull’: the positive value of living with others on terms that they could not reasonably reject.⁶⁰

From the strength of the sense of guilt, we can further understand how important the relation of mutual recognition is in our practical reasoning. Due to the importance of this relationship, reasonableness can have absolute priority over other values, including rational interest. Scanlon argues that nearly no reasons can override the consideration of reasonableness. ‘The fact that an action would be wrong constitutes sufficient reason not to do it (almost) no matter what other considerations there might be in its favor’.⁶¹ Therefore, if rational interests conflict with the constraint of justifiability, they ought to put their rational interests aside. If they refuse to do so, then it is merely because they ignore the weight of mutual respect in their practical reason. The key to practical reason is to guarantee that the reason for an action can be justified to others, rather than to pursue interests effectively. That is why Scanlon says,

[Thinking about what principle is justified is] not about what would be most likely to advance their interests or to produce agreement in their actual circumstances or in any more idealized situation, but rather a judgment about the suitability of certain principles to serve as a basis of mutual recognition and accommodation.⁶²

5.3.2 Is the conception of reasonableness neutral?

Critics might doubt that these phenomenological descriptions are enough to prove the generality of the conception of reasonableness. They might accept Scanlon’s phenomenological description, but doubt how general this description is. As Scanlon acknowledges, reasonableness as a conception of practical reason, is not wholly

⁶⁰ Scanlon (1998: 162).

⁶¹ Scanlon (1998: 148).

⁶² Scanlon (1998: 194).

neutral since it presupposes an aim of ‘living with other reasonable people’.⁶³ It means that if a person does not share this aim, then he or she has no reason to take the constraint of justifiability seriously. The phenomenological description does not apply to those who do not share this aim.

Thus critics may argue that the conception of reasonableness is less acceptable than the conception of rationality, for this conception presupposes a shared goal, but the conception of rationality does not.⁶⁴ Unlike the conception of rationality, which is always misunderstood as presupposing egoism or hedonism, it is hard to deny that reasonableness presupposes *something*, since reasonableness represents a process to test the public justifiability of a desire. If a person lacked any interest in public justifiability, then this process would be meaningless. However, this point should not be exaggerated because even if this point is correct, it does not mean that the conception of reasonableness is not widely accepted. Although the conception of reasonableness presupposes a concern with justifiability, this concern is an uncontroversial concern that people naturally accept. It seems so natural that, if one wants to cooperate with another, then one has to care about the claim of the other and cannot ignore the justifiability of one’s actions.⁶⁵ Since cooperating with one another is unavoidable, a concern with justifiability is unavoidable. Moreover, the phenomenological stance that Scanlon arrives at is also sound to most people. It is hard to deny that a sense of guilt would rise if we found that we had done something unjustifiable to others. Since this phenomenon is common, this suggests that reasonableness is a widely shared conception of practical reason even though it presupposes a goal.

⁶³ Scanlon (1998: 154).

⁶⁴ It is misleading to say rationality does not presuppose *any* goal, for in the last chapter we saw that the conception of rationality presupposes that all preferences are self-regarding, which is why it fails to capture the other-regarding motivation to be reasonable. Nevertheless, since this definition of self-regardingness is very loose, we can still say that the conception of rationality *comes close to* not presupposing any goal.

⁶⁵ Darwall (2009: 13-21).

5.3.3 The empirical evidence of general acceptability

The generality of the conception of reasonableness can be further proved by neuroscientific experiments. In such experiments, psychologists show that people naturally care about their behaviours being justifiable to others.⁶⁶ When people find that their behaviour would be unjustified to others, they usually refrain from doing it. But this inclination is sometimes inexplicit only because it is ‘diluted’ by the distance between perpetrators and victims. It will become more explicit when one is facing those to whom one needs to justify oneself in an ‘up-close-and-personal’ manner.

For example, if people receive an advertisement from an international aid organization, they will not care so much even if they know that ignoring this advertisement may not be justifiable to those poor people in other parts of the world. However, if people encounter a bleeding stranger face to face when they are driving, then they will probably feel bad if they leave this stranger alone. People are less sensitive to the unjustifiability of their actions in the former case. That is why sometimes we wrongly think that justifiability does not occupy an important role in our decision-making. But in the latter example, most of the people would be uncomfortable with their actions if they passed by, and will therefore not ignore the sufferer. This inclination is actually so common that it can hardly be seen as a cultural product. No matter which culture one was born in to, one is still concerned that one’s actions are justifiable to others. These empirical experiments show that the conception of reasonableness is a conception of practical reason generally accepted by the majority of people.

In conclusion, Scanlon believes that the conception of reasonableness is widely acceptable to actual people because actual people usually take reasonableness as overridingly important when engaged in practical reasoning. The absolute priority of

⁶⁶ Greene (2003: 847-850).

reasonableness can be seen in the strong sense of guilt from which people suffer when they harm the relationship of mutual recognition. In order to keep this valuable relationship, actual people always ‘press for the advancement of their conception of the good within reasonable limits’.⁶⁷ Hence the Kantian contractarian model based on the conception of reasonableness can satisfy the condition of generality. In the next section, however I will show that the conception of rationality actually plays an important role in our practical reason too, and Kantian contractarianism fails to fulfil the condition of priority because it overlooks the importance of rationality.

5.4 Rational rejectability of Kantian contractarianism

In the last chapter, we found that the normative force of Hobbesian contractarianism is inadequate because it overlooks the fact that, apart from rationality, reasonableness is also one of the aspects of practical reason. Kantian contractarianism is better at this point since it does not overlook the existence of rationality, yet it doubts whether rationality occupies any crucial role in the justification of political principles. This section is an attempt to show that Kantian contractarianism underestimates the importance of rationality. In fact, reasonableness could not enjoy such an absolute priority in practical reason. The importance of rationality casts doubt on how Kantian contractarianism could fulfil the condition of priority.

5.4.1 The conception of rationality in Kantian contractarianism

It should first be noted that Scanlon uses the term ‘rationality’ in a different sense from the ordinary usage. Throughout this dissertation, the term ‘rationality’ usually refers to the *instrumental* conception of rationality, which is taken to mean what most conduces to the fulfilment of the agent’s aims (and lies at the core of the conception of practical reason of Hobbesian contractarianism). Scanlon believes that

⁶⁷ Barry (1995a: 108).

‘this conception of rationality is mistaken’.⁶⁸ He defines rationality, instead, by arguing that being rational is to do ‘what we have most reason to do’.⁶⁹ In the context of his discussion, rationality refers to standing in a relationship of mutual recognition, for the great value of this relationship constitutes the (near) strongest reason for people to act. Therefore, to Scanlon, both rationality and reasonableness demand that people be motivated by the strongest reason. The only difference is that the former is harsher than the latter. Rationality requires people to think thoroughly, get as much information as possible, and ensure that they choose the strongest reason. Reasonableness, however, merely requires people to choose in ‘relative to a specific body or information and a specific range of reasons, both of which may be less than complete’.⁷⁰ ‘In between the minimum standards marked out by the idea of irrationality and the ideal of what it would be the (most) rational to believe or do, there are the notions of what is reasonable and unreasonable’.⁷¹ Under this definition of rationality, undoubtedly there are no conflicts between rationality and reasonableness, since both of them appeal to the same criteria. The problem with Scanlon’s definition of rationality will be discussed in Section 5.5.2. I would like only to point out that, at least when rationality is understood in an instrumental sense, Scanlon would not deny that there may be conflicts between rationality and reasonableness.⁷² But when these conflicts arise, Scanlon believes that reasonableness is undoubtedly more important than rationality.⁷³

⁶⁸ Scanlon (1998:192).

⁶⁹ Scanlon (1998: 30).

⁷⁰ Scanlon (1998: 32).

⁷¹ Scanlon (1998: 32).

⁷² Scanlon (1998: 192-193). Scanlon also claims that his theory is not incompatible with rational choice theory, since rational choice theory concerns the structure which the preferences of a rational individual will have (whatever the content or ground of these preferences may be), while Scanlon’s theory investigates the reasons people have for acting. (Scanlon 1998: 116) However, it seems hard to deny that rational choice theorists also have presuppositions about reasons for action. Scanlon is correct that rational choice theorists do not presuppose that people always have a reason to maximize their utility, but rational choice theorists do presuppose that people are always motivated by *considered* and *coherent* preferences (This is also why, in Hobbesian contractarianism, actual people would be motivated to follow the decisions of hypothetical contractors, for the preferences of hypothetical contractors represent the considered and coherent preferences of actual people.) Given that these preferences are not ill-formed, they will be taken as a factor of this person’s well-being and

As W. M. Sibley argues, the distinction between rationality and reasonableness is not only a technical matter, but also a familiar distinction in ordinary language.⁷⁴ Rationality usually means that people have some considered ends and have a rational plan for realizing these ends in order to maximize their utility, whereas reasonableness usually refers to following principles which could be agreed in a situation in which everyone would respect the constraint of justifiability. The conflicts between these two standards are common and obvious, since the most effective way to pursue one's own ends usually involves acts that could not be justifiable to others. When these two standards conflict, Scanlon requires people to sacrifice their rational plans for the sake of behaving reasonably. In general, people have two ways of doing so. Firstly, persons keep the considered ends of their rational plans, but achieve them by means which are compatible within the constraint of justifiability. Another way is to give up the considered ends of their rational plan and pursue some other ends which are compatible within the constraint of justifiability. However, both of these two ways go against rationality: one is asking people to be ineffective, and the other is asking people to be self-contradictory. I would like to argue that either of these two ways would create a sense of loss which is not weaker than the sense of guilt from which people would suffer in an unreasonable situation.

it is rational for this person to satisfy them effectively. Nevertheless, these preferences may still be incompatible with the constraint of justifiability. So the conflict between rationality and reasonableness still arises. For the relationship between well-being and informed preferences, see Wolff (2004: 43-44).

⁷³ Scanlon (1998: 194).

⁷⁴ Sibley (1953).

5.4.2 Why Kantian contractarianism is to be rationally rejected? (1): the ineffectiveness critique

In the first way, Scanlon might argue that our rational plan consists of ends of life and the most effective means to achieve these ends. People should keep their desired ends but pursue them by means which are compatible within the constraint of justifiability.⁷⁵ Nevertheless, I doubt that this way works. This is because this requirement underestimates our inclination to behave effectively. Effectiveness is one of the fundamental criteria in rationality. It means that, given that we have certain *ends*, we are warranted in adopting the means which can fulfil the ends to the fullest possible extent with the least expenditure. Its binding power is unconditional. As G. H. von Wright said, 'its binding nature stems from the fact that its conclusion declares the intention which an agent is logically bound to have within the teleological frame which in the premises he acknowledges for his prospective action'.⁷⁶ Suppose that I want to be a football player, I should train hard and look for chances to play in top football clubs. The principle of effectiveness is uncontroversial because it represents an internal consistency in rational deliberation. As we value our ends highly, we surely prefer taking the shortest route to realizing them. If we take an alternative, a longer route, then this action contradicts our desire to realize certain ends and we act in an inconsistent way.

If we find that we have used an ineffective means, we normally think that we did a silly thing. A sense of loss might arise, since we have behaved incoherently. We intend to give up that ineffective means immediately and rectify the error. From this sense of loss one can see the weight of effectiveness in our practical reason. Scanlon might respond that, although failing to be effective causes a sense of loss, this sense

⁷⁵ Actually it is also an approach which is explicated adopted by some Kantian contractarians, such as Brian Barry. Barry rarely requests people to give up their ends of life. He only argues that people should achieve their ends of life in a way which could not be reasonably rejected. See Barry (1995a: 82-86).

⁷⁶ Wright (1983: 59).

of loss is not that strong because most people reason irrationally on many simple tasks.⁷⁷ For example, we do not always go from store to store to find the cheapest price. Even if we find that our purchase is sold cheaper elsewhere, we do not always think of it as important. But it does not mean that effectiveness does not play an important role in our practical reason. We do not care about effectiveness in this instance simply because it is too minor. If the end is something fundamentally important in our life after we have thoroughly reflected upon it, then we usually want to pursue it as effectively as we can.

The sense of loss which we suffer from when we are forced to give up effective means of achieving our ends in life is so strong that sometimes we may even prefer being unreasonable for the sake of avoiding this sense of loss. Suppose people clearly know their ends of life and how to achieve them effectively, but then are forced to take a longer and less effective route to realize them. That is to say, they are forced to act incoherently on a matter which is most important to them. The sense of dissatisfaction created may be difficult to overcome. Even though harming the relationship of mutual recognition may bring a strong sense of guilt, this sense of guilt does not necessarily override the sense of loss caused by behaving ineffectively. As Marcia Baron points out, it is not uncommon and incomprehensible for people to behave unjustifiably to others for the sake of protecting something important in their life. 'We think well of them for being tempted to do so'.⁷⁸ Therefore, while Scanlon argues that we should always give reasonableness priority over other considerations because by doing so we can avoid the sense of guilt, we should not ignore the fact that a sense of loss, which is similarly strong, will also be created if people are forced to behave ineffectively in pursuing their ends of life.

⁷⁷ Examples of irrational thinking in the ordinary life are in Ludwig (2004: 348-349) and Wolff (1977: 138-139).

⁷⁸ Baron (1991: 855-856).

5.4.3 Why Kantian contractarianism is to be rationally rejected? (2): the self-contradiction critique

In light of the importance of effectiveness in our rationality, the first way is unlikely to work. Nevertheless, Scanlon might argue that people should make concessions in a second way, that is, they should give up their desirable ends and pursue some others which are compatible with reasonableness. Reasonableness presupposes an end of ‘living with others who are also reasonable’. When the rational plans of people conflict with reasonableness, people should recognize the weight of this and give up other ends which are incompatible. If not, we will no longer stand in a relation of mutual recognition with one another, which is a highly desirable relationship.

However, what is the reason for people to give up their existing ends for the sake of ‘living with others who are also reasonable’? It must presuppose that people prefer the end of ‘living with others who are also reasonable’ to their existing ends. In the last chapter, we saw that rational people reflect thoroughly upon their ends and rank their ends in terms of utility. The ends that they want to achieve more have higher utility, while the ends that they want to achieve less have lower utility.⁷⁹ Given that time and resources are limited, a rational person prefers pursuing choices with higher utility to choices with lower utility. If one believes that she prefers A to B but still gives up A for the sake of B, then his behaviour becomes self-contradictory. This is because, on one hand, she desires A more than B, but, on the other hand, she pursues B instead of A. His belief is then contradictory to his action.

But it is unlikely that the end of ‘living with others who are also reasonable’ is always preferable to other ends. Rationality merely requires people to reflect upon their ends in order to ensure that they are *considered* and *coherent*. But there is no

⁷⁹ Bicchieri (2004: 183-185).

guarantee that their considered and coherent ends must include the end of ‘living with others who are also reasonable’ (or ends compatible with this end). After reflection, people find ends which are most valuable to them, and these ends may be incompatible with the end of ‘living with others who are also reasonable’. These ends give meaning to one’s life to the extent that one may feel one’s life has nothing left if one is deprived of these ends. Compared with losing these ends of life, I cannot see that damaging the relationship of mutual recognition can bring the same degree of suffering. As Jay Wallace argues, ‘how, we may wonder, can the abstract goods realized by compliance with moral principles possibly prevail in competition with the contingent ambitions and relationships that give meaning and texture to our lives?’⁸⁰ The relationship of mutual recognition is important. We value it and we would feel guilty if this relationship were harmed. Nevertheless, it is usually only one of the desirable things in our life. It cannot be compared with our considered ends of life, which are of irreplaceably valuable.⁸¹

If people choose to behave reasonably instead of rationally pursuing their considered ends, then they act in a way that runs counter to what they believe. Since the pursuit of ends of life is one of the most important matters in one’s life, one can hardly accept self-contradiction in this issue. If Scanlon justifies the priority of reasonableness by the pain of guilt that is aroused when we harm the relationship of mutual recognition, then this justification is unpersuasive because we feel far more pain when we give up our ends of life. Asking one to change one’s ends of life for the sake of being reasonable is less acceptable than Scanlon thinks, because it requires one to act in a contradictory fashion in something which is the most significant in one’s life.

⁸⁰ Wallace (2002: 454).

⁸¹ Some might argue that this presupposes a desire-based of reasons which Scanlon rejects. However, in Scanlon (1982), the earlier Scanlon in fact presupposes a desire-based view of reasons. Although he rejects this view in Scanlon (1998) and endorses a value-based, externalist view of reasons, this brings him other more serious problems and I will discuss this issue in Section 5.6.3.

In these two sections, I discussed two cases where people are forced to act irrationally for the sake of reasonableness. First, people may have to give up their *ends* of life for the sake of being reasonable, but then they would act in a self-contradictory manner. Secondly, people may have to use a less effective means to pursue their ends of life, but then they would act ineffectively. In both cases, a strong sense of loss is aroused. Since the strength of this sense of loss is not weaker than the sense of guilt created by unreasonableness, Scanlon could no longer argue that one should always give reasonableness absolute priority in order to avoid guilt. This strong sense of loss also shows that reasonableness does not, as Scanlon argues, have an overriding status over rationality. In practical reasoning, rationality is not any less important than reasonableness. While reasonableness justifies obligations to behave in a way which are justifiable to each other, rationality justifies obligations to pursue ends in the most effective way. People have no necessity of taking the former obligations as more important than the later obligations.

Given the equal importance of rationality and reasonableness, Kantian contractarianism, which is based on the conception of reasonableness only, fails to satisfy the condition of priority. The problem of Kantian contractarianism is that it underestimates the importance of rationality in practical reason. Although Scanlon shows that his contractarian principles are reasonably justified, he does not show that they are rationally justified as well. Rationality may justify obligations that are incompatible with these contractarian principles. Thus people who are governed by these principles will have two conflicting obligations, neither of which can override the other. Kantian contractarianism fails to justify political principles which have an overriding authority. It should not overlook the fact that, while people naturally want to be reasonable, they naturally want to be rational as well.

5.5 Should reasonableness be considered prior? (1): two arguments

The last section shows that the absolute priority of reasonableness implies that people are forced either to be ineffective or to be self-contradictory, which are both

irrational behaviours. In order to satisfy the condition of priority, Scanlon has to explain why people should concern themselves about reasonableness over rationality. Scanlon might acknowledge that people should pursue ends rationally but argues that people should choose proper ends.⁸² It is rational for people to give up those desirable values that conflict with reasonableness and choose other values to pursue. So Scanlon has to explain why people have to prefer being reasonable to pursuing those values that they highly desire in their lives even after reflection. He has a 'three-part' strategy to defend this thesis. 'It can make sense, if we recognize values other than [reasonableness] and take them seriously, to claim that reasons of [being reasonable] have priority over all the rest'.⁸³

The first is to argue that insofar as these are things that people have reason to pursue and to value, these reasons will be among those that can make it reasonable to reject some principles. Therefore there will be pressure within the morality of right and wrong to make room for these values. But there will of course be limits, and the second part of the strategy (which divides into two subparts) is to argue that when these limits are reached we have good reason to give priority to the demands of right and wrong. This can be done in part by appealing to the great importance of justifiability to others and to the particular interests that moral principles protect, and in part by arguing that the other values, properly understood, have a built-in sensitivity to the demands of right and wrong.⁸⁴

I will discuss the first two strategies in section 5.5 and discuss the third strategy in section 5.6.

⁸² Scanlon's acknowledgement on the importance of effectiveness can be seen in Scanlon (1975: 75-76).

⁸³ Scanlon (1998: 148).

⁸⁴ Scanlon (1998: 166).

5.5.1 The leave-room argument

The first strategy which Scanlon suggests is to justify the priority of reasonableness by showing that, since reasonableness does leave enough room for other values, those other values should give way to reasonableness, should they conflict. Reasonableness leaves room for these values because ‘insofar as these values are things that people have reason to pursue and to value, these reasons will be among those that can make it reasonable to reject some principles’.⁸⁵ That is to say, these values have already been taken into account when we consider whether a principle could be reasonably rejected. ‘An individual will thus have a [generic reason] for wanting to reject a principle if the results of its general acceptance would be very bad from the point of view of that person’s conception of substantive good’.⁸⁶ If political principles always require people to make sacrifices in their rational pursuits of values, then this principle could be reasonably rejected. Hence, justified political principles have already provided ‘an acceptable distribution of control over important factors in our lives’.⁸⁷

On the other hand, if some values are eventually rejected by a reasonable agreement even after considering those generic reasons, we cannot oppose the demand of reasonableness in the name of these values. The importance of these values has its limits. This is only one of the generic reasons, and it can be overridden by other generic reasons. If the rational pursuit of some values would impose severe burdens on other people, then this would be unjustifiable to other people. In this situation, these values should be given up since they could be reasonably rejected. Nevertheless, it does not mean that people have no values to pursue, for apart from those values that conflict with reasonableness, they could still choose many other values as their *ends*, insofar as these values could not be reasonably rejected. In

⁸⁵ Scanlon (1998: 161).

⁸⁶ Scanlon (1993: 183).

⁸⁷ Scanlon (1978: 37).

short, Scanlon argues that reasonableness does give adequate attention to the importance of rationality. The constraint of justifiability already allows plenty of values for people to rationally pursue. They are then justified in prohibiting others from pursuing them.

To me, this argument wrongly conceives the nature of values that are taken as *ends* of life in rational plans after thorough reflection. It seems to take rationally planned values to be something which could easily be replaced by other values. For example, I want to buy an apple and there are no apples in the supermarket, but there is still a wide variety of other fruits; perhaps I should buy an orange. However, the importance of an apple cannot be comparable to the importance of these values. Suppose that my dream is to be a football player, I would not give up my plan to be a football player simply because there are many other careers available. These values are usually nearly unique to the extent that they are difficult to replace, and this is why they are taken as *ends of life*. They are closely connected to people's lives to the extent that the meaningfulness of their life would be diminished if they gave up *their* values. Although there are many other values available, these values are nothing more than irrelevant choices. Scanlon is right that people have a wide variety of choices, but most of these choices may not interest them. People will not give up their rational plans of life simply for the reason that other rational plans would still be available if they gave up their plan.

In fact, this argument is misleading in the justification for the priority of reasonableness. The point in question should be whether the aim of 'living with others who are also reasonable', which is the aim presupposed in reasonableness, can override the other values in people's rational plan of life. Only when the former can override the latter, would people have reason to take the conception of reasonableness as having the highest priority and would therefore arguably give up other values that conflict with it. This ought to be what Scanlon has to prove. Merely showing that the conception of reasonableness already takes other values into account and leaves plenty of values for people to choose is not enough. For even if

this proves true, provided that people take other values as being more important than the aim of ‘living with others who are also reasonable’, they still have no reason to prefer behaving reasonably over pursuing other values. Hence, in the question of whether reasonableness can override rationality, the key point is the comparison between the aim of ‘living with others who are also reasonable’ and ‘other’ values. The ‘leave-room’ argument simply avoids this key point.

5.5.2 The proper-understanding argument

So we can move to Scanlon’s second strategy; the proper-understanding argument. Scanlon could argue that the conflict between rationality and reasonableness is based on a wrong understanding of rationality. As I noted, Scanlon has his own definition of rationality. He disagrees with the instrumental conception of rationality and defines it in a more substantial way.⁸⁸ In his definition, rationality and reasonableness are nearly the same criteria. They are only different to the extent that the former requires people to think with a full set of information, but the latter merely requires people to think with ‘a specific body of information...less than complete’.⁸⁹ Both require people to pursue ends that are supported by good reasons which could also be recognized by others, only that ‘reasonableness uses the idea of the most rational thing to do or think in a more realistic sense’.⁹⁰ If others could suggest stronger reasons (such as the value of the relationship of mutual recognition, which is the (nearly) strongest reason) which can explain why certain ends should be rejected, then this rational person should give up these ends. Hence, Scanlon could argue that if we understand rationality properly, we would find that rationality and reasonableness both demand we act according to strong reasons, and there would therefore be no conflict.

⁸⁸ Scanlon (1998: 192-193).

⁸⁹ Scanlon (1998: 32).

⁹⁰ Marshall (2003: 16).

Whether people always think the value of the relationship of mutual recognition is the strongest reason, will be discussed in the next section. I would like to discuss here whether the conflict between rationality and reasonableness can be avoided by reinterpreting the conception of rationality. Scanlon's reinterpretation of rationality is problematic since it goes against our ordinary understanding of rationality.⁹¹ In our ordinary understanding, rationality and reasonableness are usually two separate norms of practical reason which are distinctively different from each other. Rationality is more about 'interest' and how to pursue these interests effectively, whereas reasonableness is more about 'constraint' and whether the pursuit of interests violates equal relationships among people. Reasonableness is more seen as a *social* virtue. It involves 'a willingness to listen to the reasons offered by others amounts to an openness to their perspectives and interests.'⁹² But rationality is 'essentially an *intellectual* virtue,' which merely requires people to ascertain their preferred ends, and to reach these ends by the best means.⁹³ Hence, we usually think that a rational person can be unreasonable.⁹⁴ For example, a killer constructs a murderous plan very rationally, but what this killer is doing is unreasonable. The reinterpretation of Scanlon seems to go against this intuition.

A more serious problem with this strategy is that Scanlon in the end does not explain why his interpretation is a better interpretation of rationality. While Scanlon can interpret rationality in a way that is compatible with reasonableness, it is always open to opponents to offer alternative interpretations in order to show that rationality *indeed* conflicts with reasonableness. So why ought Scanlon's interpretation to be accepted over other alternatives? Scanlon might insist that other interpretations would be proved wrong if a 'proper understanding' were reached. Yet this is merely begging the question, since he does not explain clearly why his interpretation is

⁹¹ A similar point is suggested by Susan Mendus. Although she argues about the conflict between friendship and morality, she also believes that this strategy of reinterpretation distorts our ordinary understanding of those values which genuinely conflict with morality. See Mendus (2003).

⁹² Miller (1987: 420).

⁹³ Sibley (1953: 555-556).

⁹⁴ This observation is also suggested by Samuel Scheffler. See Scheffler (1992: 52-54).

‘proper’. If an interpretation is ‘proper’ only when it is compatible with the priority of reasonableness, then the proper-understanding argument is circular. Scanlon cannot escape from the responsibility for explaining and providing evidence in support of the claim that his interpretation is more ‘proper’.

5.6 Should reasonableness be considered prior? (2): the great-value argument

The last argument which Scanlon could use to justify the priority of reasonableness is the ‘great-value’ argument, which justifies the priority of reasonableness by the great value of reasonableness in our life and explains why we should consider reasonableness before other values. This argument is the strongest in Scanlon’s ‘three-part’ strategy because it explains why the end of ‘living with others who are also reasonable’ which is the aim presupposes in reasonableness, can override other valuable ends.⁹⁵ In light of Scanlon’s externalist view of reason, achieving this end means that people can enjoy certain valuable relationships while satisfying certain interests. A person can make her life more valuable ‘by making herself a morally better person’.⁹⁶ Hence, no matter what ends people value, reasonableness can still override these ends. In section 5.6, I will explain why this argument cannot hold, and, although reasonableness is valuable, it does not necessarily mean that actual people should give up their desirable ends of life for the sake of reasonableness.

⁹⁵ Although this point is implicit in *What We Owe to Each Other*, it is made more explicit in Scanlon’s another article, ‘Value, Desire, and Quality of Life’. In this article, when Scanlon explains why people should give up the pursuits of their conception of substantial good for the sake of principles that no one could reasonably reject, he usually refers to the ‘moral aim of living with others who are also reasonable’. See Scanlon (1993: 185-186).

⁹⁶ Scanlon (1993: 170).

5.6.1 The great-value argument

According to Scanlon, the valuableness of reasonableness is explained by two reasons: one is the ‘great importance of justifiability to others’, the other is the ‘particular interests that moral principles protect’.⁹⁷ The importance of justifiability comes from the importance of the relation of mutual recognition, which is a relationship ‘appealing in itself—worth seeking for its own sake’.⁹⁸ The relationship represents a desirable moral ideal, a ‘Kingdom of Ends on the cheap’.⁹⁹ Living in such a moral community is a state of affairs that is desired by everyone. The value of this relationship can also be understood in a negative way, that is, in a way which shows the feeling of being unwelcome and guilt that appears when we harm this relationship. Scanlon argues that, if one finds that one’s action is unjustifiable to others, then a sense of guilt is aroused. This sense of guilt comes from our recognition of the fact that we have violated a valuable relationship with others. Since this sense of guilt is quite strong, it implies that mutual recognition occupies an important role in our practical reasoning.

Apart from the desirability of the relation of mutual recognition, Scanlon also thinks that being reasonable can bring us pleasure and satisfy some of our particular preferences, although this point seldom appears in his writings. Scanlon argues that, apart from the sense of guilt which appears when we behave unreasonably, there is also a loss of pleasure, which ‘is not merely a matter of feeling guilty or distressed at the thought that one’s life and institutions do not measure up to one’s moral goals’.¹⁰⁰ Hence,

one cannot take the same pleasure in one’s cooperative relations with others as members of the same firm or university, say, if one comes to believe that they are being asked to participate on

⁹⁷ Scanlon (1998: 166).

⁹⁸ Scanlon (1998: 162).

⁹⁹ Scanlon (1998: 162). Scanlon (2009: 189-190).

¹⁰⁰ Scanlon (1998: 163).

terms they could reasonably reject, and the meaning of one's own successes and accomplishments is undermined by the thought that they were attained on terms that were basically unfair.¹⁰¹

Accordingly, reasonableness is valuable in two dimensions: one concerns relations and the other well-being. The first is the value of the relation of mutual recognition. This relation is valuable because it represents a 'Kingdom of ends on the cheap', which is an appealing state of affairs. The second one is the pleasure that they can enjoy as long as they participate in fair terms of cooperation with others. Although Scanlon does not explain in detail where the pleasure comes from, he seems to believe that participating in fair terms of cooperation with people around us represents a kind of success and accomplishment. This success and accomplishment brings satisfaction to the members of cooperation. By considering these two values, Scanlon contends that people will be willing to consider reasonableness prior to other values when they conflict with each other.

5.6.2 Valuableness and priority

Undoubtedly Scanlon's account can explain why reasonableness is valuable and we should take this into account. However, his account is unable to explain why reasonableness has an *absolute* priority over any other values. I will show that, at the very least, Scanlon's account cannot explain why people should always think that reasonableness is more valuable than other ends that they rationally desire. I will first discuss the value of the relation of mutual recognition. No one would deny that the 'Kingdom of ends on the cheap' is a desirable moral ideal. As Scanlon argues, being a reasonable person makes a person's life more valuable.¹⁰² However, this is only one of the many ways of improving one's quality of life. A person may favour the

¹⁰¹ Scanlon (1998: 163-164).

¹⁰² This is also acknowledged by Scanlon, who remarks that, apart from being morally better, a person's life could also be improved by 'aiming at other things that she takes to be worthwhile'. See Scanlon (1993: 170, 175).

‘Kingdom of ends on the cheap,’ but not take it as the most desirable ideal even after thorough rational deliberation. There are other ideals that people may favour. In this instance, if they follow the demand of reasonableness, they have to change or drop those values that they rationally desire most. But why should they do so? Scanlon could answer this question by referring to the great value of the ‘Kingdom of ends on the cheap’ again. Yet this is not an adequate response, for, in the mind of these people, the ‘Kingdom of ends on the cheap’ is at most the second most desirable value. Why should they give up their most desirable value, for the sake of achieving the ‘Kingdom of ends on the cheap’? Scanlon has to provide an alternative reason for explaining why the ‘Kingdom of ends on the cheap’ can justify the absolute priority of reasonableness. If not, he is begging the question. He cannot explain why one should give up the ideal which one values most and pursue the ‘Kingdom of ends on the cheap’ simply by referring to the great value of the ‘Kingdom of ends on the cheap’.

This alternative reason could be the second dimension of the value of reasonableness, that is, one might gain pleasure from being in a fair scheme of cooperation. If this pleasure is a kind of *summum bonum*, the highest good which human beings always wants to achieve, then we can understand why people should always consider reasonableness prior to other things, for no other things can bring greater pleasure than reasonableness. Yet this is not so. We can easily find cases in which people achieve greater pleasure from activities apart from participating in a fair scheme of cooperation. Human beings differ in their understanding of pleasure. Some religious people may take salvation to be the highest good and, in order to achieve salvation, strictly follow their religious teachings. To these people, the pleasure achieved in participating in a fair scheme of cooperation is just not worth considering when compared with the pleasure they will enjoy in Heaven. So why should they sacrifice the pleasure of salvation for the pleasure of participating in a fair scheme of cooperation? This kind of case is not uncommon, for, in fact, people rarely take the pleasure which they could enjoy through participating in a fair scheme of cooperation as an overriding good in their life. Participating in a fair scheme of cooperation is merely one of the sources of pleasure in one’s life. Therefore,

although Scanlon is right that following the demand of reasonableness can bring us pleasure, it is doubtful that this reason is strong enough to push people to consider reasonableness to have priority over rationality.

5.6.3 The externalist account and its costs

Scanlon could further argue that the overriding value of reasonableness cannot be fully understood without referring to the ‘universality of reason judgments’.¹⁰³ Unlike *internalists*, who believe the reasoning behind why people have reason to act, depends on what motivations they have, Scanlon holds an *externalist* account and believes that reason should have a more independent status. Given that two persons are in a similar situation, a reason which can move the first person should be also able to move the second person, no matter what motivations they both have:

The universality of reason judgments is a formal consequence of the fact that taking something to be a reason for acting is not a mere pro-attitude toward some action, but rather a judgment that takes certain considerations as sufficient grounds for its conclusion. Whenever we make judgments about our own reasons, we are committed to claims about the reasons that other people have, or would have under certain circumstances.¹⁰⁴

Insofar as a reason has sufficient grounding, everyone should accept it regardless of what desirable ends it has. Based on this externalist account, Scanlon could argue that the value of reasonableness is actually an external reason, which could move actual people regardless of their motivations. Some people might wrongly estimate the real weight of this reason and therefore might wrongly think that other ends in

¹⁰³ This possibility is suggested by Sarah Marshall, who argues that the inescapability of moral requirement in Scanlon’s contract theory cannot be understood without the principle of the universality of reason judgments. See Marshall (2003: 30).

¹⁰⁴ Scanlon (1998: 74).

their life are more valuable than reasonableness.¹⁰⁵ However, the fact that some people fail to be moved by this reason so much does not imply that this is not a strong reason. Rather, Scanlon believes that these people are merely exceptional cases. We should question why these people fail to recognize the weight of the reason, but not doubt the reason itself. At the end, even if people fail to recognize the overriding value of the relationship of mutual recognition, still, they *should* consider reasonableness prior, for the objective weight of this reason is not affected by their ‘strength of subjective preferences’.¹⁰⁶

If [a conception] that I favour leaves you cold, you may not have reason to adopt it. But if it is a worthwhile conception then you do have reason not to scorn it and reason not to mock those who take it seriously. If you fail to see that you have such reasons, and would still fail to see this even after the most complete process of imaginative reflection you could manage, this indicates a kind of deficiency on your part—moral narrow-mindedness, we might call it.¹⁰⁷

Scanlon further uses the maltreatment of a wife as an example.¹⁰⁸ He argues that a husband should be able to recognize the reason for not treating his wife badly. If he fails to see this reason, then it is his deficiency. His cruelty does not affect the strength of this reason. This externalist view of reason can explain why actual people should appreciate the valuableness of reasonableness and give it absolute priority.

There are surely some people who take torturing others as nothing or endorse deeply repugnant views such as Nazi. Since these people are too rare, we usually think that these people are merely exceptional cases. Although their views may be

¹⁰⁵ According to Scanlon’s theory of valuableness and well-being, whether something is valuable and can make a person’s life better depends upon whether this thing is supported by good reasons. Hence, if the ends in a person’s life are incompatible with reasonableness, then it means that these ends are unjustified. So these ends cannot be good reasons, thus cannot be valuable. It is a misunderstanding to think that these ends in life are more valuable than reasonableness. See Scanlon (1998: 89-90, 118-123).

¹⁰⁶ Scanlon (1975: 73).

¹⁰⁷ Scanlon (1998: 370).

¹⁰⁸ Scanlon (1998: 369-370).

instrumentally rational, we may still take them as being mentally deficient and reject their views. However, these cases are in fact misleading examples because they are too uncontroversial. We should not confuse these extreme cases with the controversial case about the overriding value of reasonableness. To many people, the value of reasonableness is merely desirable, but not overridingly desirable. They may have other ideals or relationships that are much more valuable from their perspectives. Compared to the case of maltreating a wife, it is much more controversial for Scanlon to claim that everyone *should* take reasonableness as an overriding value. As Gerald Dworkin argues, in many cases, Scanlon is too dogmatic in declaring that those who fail to be motivated as Scanlon describes are ‘deficient’:

If there is no compelling argument against the view that [those who are not motivated appropriately in Scanlon’s sense] have no reason to act morally (because they do not share our ideal of reaching general agreement, and could not be brought to share it by any mode of reasoning or further experience), then we cannot say that such a person has a reason to not act immorally, only that they are deficient in not seeing that there is such a reason.¹⁰⁹

If Scanlon insists that the priority of reasonableness is based on external reasons that everyone should accept, then his theory will lose several virtues of contractarianism. One of the virtues of contractarianism is that it shows respect for the will and choices of rational individuals. A contractarian usually disagrees that a political principle should be an arbitrary authority imposed on people.¹¹⁰ A justified political principle should show that it represents a disposition of actual individual wills. This can be shown by demonstrating that, in certain ideal circumstances, people think deliberatively and would finally choose this principle. However, the externalist view of reason is inconsistent in this respect, since the externalist view proposes that something is impermissible no matter what the wills of actual people are.¹¹¹ Even

¹⁰⁹ Dworkin (2002: 480).

¹¹⁰ This theme is also endorsed by Scanlon. See Scanlon (1977: 69).

¹¹¹ See Scanlon (2008: 99-100). In fact, this externalist view becomes more and more dominant in Scanlon’s later writings. In *Moral Dimensions*, Scanlon even argues that an act is impermissible given that it is incompatible with the idea of rational beings as ends in themselves. This nature of the act

when they disagree with the overriding value of reasonableness ‘after the most complete and careful process of reflection and deliberation,’¹¹² the weight of this value is still unaffected. Their disagreement has no influence on the justifiability of the priority of reasonableness. It seems difficult to say that the externalist view takes will and choice seriously.¹¹³

Moreover, contractarianism has another virtue: its justification for political principles does not rely on any peculiar, non-natural properties. Contractarians do not take any political authority to be self-evident. Rather, a political principle has authority only when it represents a ‘contract’; a product of subjective wills. However, the externalist view proposes that the authority of a reason is independent of subjective will. Hence, the priority of reasonableness is always justified, since its valuableness constitutes a strong, self-evident external reason. However, where does its authority come from? Scanlon has to explain how this authority can affect and bind people even if these people, after they thoroughly reflect, do not accept this authority. Surely he could justify this by an external, objective moral authority, such as natural, independent moral laws. Nevertheless, the cost of doing so is that Scanlon then could not avoid the burden of explaining how this external moral authority is compatible with the prevalent scientific, secular worldview. Contractarianism is an appealing method of justification because it takes subjective will as the starting point of political thinking and justifies principles in a way which is compatible with this

determines its moral permissibility, no matter what people take this idea to be. Subjective will plays nearly no role in determining the moral permissibility of an action. Whether or not people take the idea that others are ends in themselves as giving them reasons for action is irrelevant to the question of whether their act is wrong. This is highly different from his beliefs in his early writings. Here, Scanlon insists that consent plays a significant role in the justification of obligations and institutions. ‘Could an autonomous individual regard the state as having, not as part of a special voluntary agreement with him but as part of its normal powers qua state, the power to put such an arrangement into effect without his consent whenever it (i.e. the legislative authority) judged that to be advisable? The answer to this question seems to me to be quite clearly no’. See Scanlon (1972: 19).

¹¹² Scanlon (1998: 369).

¹¹³ This point is also suggested by Thomas Pogge, who argues that Scanlon’s contract theory is concerned too much with what hypothetical reasonable contractors would reject, but talks too little about ‘the *actual* standpoints and rejection grounds of *real* people in the world *here and now*’. See Pogge (2001: 135-138).

worldview. Therefore, insofar as Scanlon proposes that a moral authority could be justified independent of subjective will, he deviates from contractarianism.

This section shows the flaws of Scanlon's 'great-value' argument. Scanlon could justify the priority of reasonableness by referring to the great value of reasonableness, but I have also raised important doubts about the idea that reasonableness is not only valuable but also the *most* valuable. Scanlon could further argue that the overriding value of reasonableness is an external reason, so the fact that someone does not accept it does not affect its weight. Nevertheless, given that there are many 'exceptional' cases that do not take the value of reasonableness as overriding, and Scanlon's theory will lose several virtues of contractarianism if he maintains this externalist view. In conclusion, Scanlon's 'three-part' strategy fails to give a persuasive account of why, given that rationality is as important as reasonableness in practical reason, reasonableness should always be considered before rationality.

5.7 Should reasonableness be considered prior? (3): the scepticism argument

Given that Scanlon's 'three-part' strategy fails to show that Kantian contractarianism can satisfy the condition of priority, we should then broaden our discussion and consider alternative justifications provided by other Kantian contractarians. Barry suggests that reasonableness should be considered prior to rationality because the value of ends of life in our rational plan is uncertain. This section will discuss whether Barry's argument can rescue Kantian contractarianism.

Unlike Scanlon, Barry justifies the priority of reasonableness by a 'sceptical thesis'. This affirms that disagreement exists among reasonable people on the question of which conceptions of the *good* are worthwhile and which are impoverished or

meaningless.¹¹⁴ Barry claims that it is never possible to know with certainty which conceptions of the good are better than others. His claim is grounded in the observation that, in spite of discussions about the good life that have taken place over several centuries, we are still nowhere near reaching consensus.¹¹⁵

How can the case for scepticism be made out? I am inclined to think that there is a strong a priori argument for the inherent uncertainty of all conceptions of good. Any chain of reasoning of this sort is, however, open to objection by a dogmatist that it is too speculative to overcome his powerful sense of being privy to the truth.¹¹⁶

According to the ‘sceptical thesis’, we should not take the values in our rational plan as undoubtedly meaningful. Failure to follow our rational plan of life is thus not a big problem, because we may find that it is wrong or meaningless in the future. Compared with being irrational, we are more certain that being unreasonable is the same as harming equal respect. Therefore, we should not insist on our rational plan as long as it could be reasonably rejected, for we may finally behave unreasonably for the sake of something which is actually wrong or meaningless.

Barry is right that people cannot be sure that their values in a rational plan are meaningful. However, I am not sure that people are as strongly sceptical about the values in their rational plans as Barry asserts. The sceptical thesis requires that people should always remain doubtful about their values. Yet doubt cannot arise without evidence. One doubts something only when one has discovered some evidence which shows that that one’s thinking is questionable. If not, the sense of doubt would be so weak to the extent that it could not affect our decision-making. This sense of doubt is even less significant in the case of rational deliberation about the *ends* of life. As the values that people take as the ends of life are something they have rationally reflected upon thoroughly, they are products of their most careful

¹¹⁴ Barry (1995a: 12, 27, 168-169).

¹¹⁵ Barry (1995a: 168-172).

¹¹⁶ Barry (1995a: 169).

consideration and represent the most reliable ground of their thoughts. In this case, they will not easily give up their desired values because of a sense of doubt, without any adequate evidence. Unless they face some clear evidence which tells them that their values are problematic, they will not have a sense of doubt which is strong enough to undermine their belief that the current ends of life in their rational plan are meaningful.

Even if the sceptical thesis is accepted, it does not mean that people would easily give up their values if they conflict with principles which could not be reasonably rejected. Barry argues that, as long as people admit that there are some uncertainties about these values, their values will become less valuable in their mind. However, I cannot see why this necessarily occurs. Although I may discover in the future that the values I took some years ago are not as valuable as I think now, it does not mean that these values were not at one time of extreme importance. My sense of loss generated when I give up this value will not be any less even after I know that these values might be of little worth. No matter whether I accept the sceptical thesis, this sense of loss will still be as painful. If the sceptical thesis could not undermine the sense of loss caused by irrationality, then it would provide no reason for people to consider reasonableness prior to the rational pursuit of these ends.

Furthermore, even if we accept the sceptical thesis, we need to know why it applies only to all judgments about the values in our rational plan, but not to judgments about the value of reasonableness. While we are uncertain about every value in our rational plan, we should also be uncertain about the value of reasonableness. Why assume that there is a distinction between the value of reasonableness (which can be known with certainty) and the values in a rational plan of life (which cannot)? The sceptical thesis does not state that we can resolve with certainty issues concerning the value of reasonableness, but not issues concerned with the values in our rational plan. It is compatible with a position which states that reasonableness is not so valuable as we think. If both the value of reasonableness and other values are subjected to the challenge of the sceptical thesis, then the value of reasonableness is not necessarily

more reliable than any other values. Therefore Barry has to provide an explanation; without it we have no reason to think that we should not insist on our ends of life if they could be reasonably rejected.

In conclusion, although Barry suggests the sceptical thesis to justify the priority of reasonableness to other values, this sceptical thesis is problematic. Even if we are uncertain about the values in our rational plan, we would not think that these values are less important. Moreover, if the worth of all values is doubtful, then the value of reasonableness cannot be exempted as well.

5.8 Conclusion

The main aim of this chapter has been to show that Kantian contractarianism overlooks the importance of rationality in our practical reason, thus fails to fulfil the condition of priority. Kantian contractarianism rightly observes the importance of reasonableness in our practical reason. Reasonableness, as a testing model, justifies only obligations that are compatible within the constraint of justifiability. Therefore they argue that justified political principles should represent a hypothetical agreement that would be made by reasonable contractors. However, as well as reasonableness, rationality is an aspect of practical reason. The conception of rationality justifies obligations to pursue interests effectively. Since both rationality and reasonableness occupy crucial roles in our practical reason, and the sense of loss caused by irrationality is as painful as the sense of guilt caused by unreasonableness, there is no reason for us to consider either one of them, but not the other, as having absolute priority. In short, since the conception of reasonableness is generally acceptable to actual people, Kantian contractarianism can satisfy the condition of generality. But since the conception of rationality, which is an equally important aspect of practical reason, is overlooked, Kantian contractarianism fails to satisfy the condition of priority.

In addition, I have examined Scanlon's 'three-part' strategy. Scanlon's 'three-part' strategy consists of the 'leave-room' argument, the 'proper-understanding' argument and the 'great-value' argument, and I have shown that none of them are compelling. In fact, the failure of these three arguments also sheds light on the relationship between rationality and reasonableness. The failure of the 'proper-understanding' argument shows that rationality and reasonableness are both distinctive conceptions of practical reason; the former cannot be reduced to the latter. The failure of the 'leave-room' argument and the 'great-value' argument shows that rational interest holds the same importance as reasonableness in practical reason; the latter cannot override the former even though it is highly valuable or leaves enough room for the considerations of the former. The failure of the 'three-part' strategy reveals the flaws of Kantian contractarianism as an inadequate account of practical reason.

If we combine the study of Kantian contractarianism with the study of Hobbesian contractarianism in the last chapter, we arrive at an interesting conclusion. These two models fail to provide a satisfactory justification of political principles in two opposing ways. By the end of the last chapter, I showed that Hobbesian contractarianism, which relies on the conception of rationality, ignores the conception of reasonableness in practical reason. In other words, this model considers human beings as rational animals but ignores that human beings are not *merely* rational animals. On the other hand, Kantian contractarianism fails in an opposite way. Its political principles are justified solely by the conception of reasonableness, but it fails to explain why being reasonable is always more important than pursuing rational interests. In other words, this model considers human beings to be reasonable animals but ignores that human beings are not *merely* reasonable animals. In short, the Hobbesian contract should care more about the conception of reasonableness, while the Kantian contract should care more about the conception of rationality. It seems that each of them has an element that the other side lacks, and each should therefore be supplemented by the other. The failure of these two models indicates that a contractarian should give sufficient weight to both ideas of rationality and reasonableness. This kind of contract model would be the 'hybrid contract model' which is the 'victim' of the orthodox dichotomy in the contemporary discussion of

Chapter 5 Kantian Contractarianism

contractarianism. Can this kind of contract model provide a satisfactory justification of political principles? This is the central issue to which I turn in the following chapter.

Chapter 6 Hybrid Contractarianism (1): the earlier Rawls

6.1 Introduction

In the last two chapters, I discussed two different contractarian models and showed their inadequacies. Hobbesian contractarianism is inadequate because it forsakes reasonableness; Kantian contractarianism is inadequate because it forsakes rationality. Since rationality and reasonableness are two aspects of practical reason, a satisfactory hypothetical contract has to be justified from both of these perspectives. Nevertheless, it is too early to claim the end of contractarianism, for there is still a third possible model which possesses the strengths of these two kinds of contractarianism—hybrid contractarianism.

Unlike Hobbesian and Kantian contractarianism, hybrid contractarianism is based on a dual conception of practical reason which assumes that people are both rational and reasonable.¹ Human beings would feel guilty if they did something which could not be justified to others, but they would also feel uncomfortable if they had to constantly sacrifice their rational ends. Therefore, a justified social agreement should be an agreement which is both rationally and reasonably justified. Compared with the other two kinds of contractarianism, hybrid contractarians are few in number. Rousseau is an early example. To Rousseau, citizens have two motivations; a sense of duty to treat others equally,² but also a fundamental interest in their freedom and in maintaining their personal independence.³ People's allegiance to the social compact is based on these two motivations. Hence, the social contract not only

¹ Rawls (1999a: 109-112).

² Rousseau (1979: 445-446).

³ Rousseau (1987: 150-151).

represents the principle of *right* in a civil society,⁴ but also represents the common good of citizens.⁵

Apart from Rousseau, Rawls is also a hybrid contractarian; however his model was not hybrid from the very start. The earliest version of Rawls' contractarian theory was a Hobbesian contractarian model. Like Gauthier, Rawls assumed that people were rationally self-interested parties that were uninterested in others. A justified social agreement should be a mutually advantageous agreement.⁶ Yet Rawls soon recognized the inadequacy of the instrumental conception of rationality. Thus Rawls and Gauthier at the end developed their contractarian models in different directions. Gauthier insisted on developing a contractarian model based on the instrumental conception of rationality, whereas Rawls gave up this project early and turned to adopt a dualistic conception of practical reason. This chapter will examine the hybrid contractarian model developed by Rawls and will discuss its strengths and weaknesses.

I will first offer an interpretation of Rawls' social contract that is different from the view accepted by most commentators. I will argue that Rawls' social contract should be understood as a public agreement among free and equal persons in a well-ordered society, and this agreement is a hybrid model because it is based on a dualistic conception of practical reason.⁷ Then I will examine Rawls' account of the

⁴ Rousseau (1987: 151-152).

⁵ Rousseau (1987: 150-151, 153-154) and Rousseau (1979: 80).

⁶ Rawls (1958).

⁷ Some people might disagree with this 'constructivist' interpretation of Rawls. Rawls adopts reflective equilibrium as his methodology. 'Justification rests upon the entire [moral] conception and how it fits in with and organizes our considered judgments in reflective equilibrium'. (Rawls 1999a: 507) In reflective equilibrium, all beliefs are provisional and can be revised in order to be organized into a coherent picture. So Rawls should be understood as a *coherentist* rather than a constructivist. However, as David Brink points out, Rawls actually has a strong constructivist character and takes a certain conception of the person for granted. Rawls assumes that people are by nature free and equal

congruence between rationality and reasonableness. Rawls argues that principles of justice do not simply make a ‘cold moral demand’, but also represent the most rational way to realize people’s fundamental goods. This can explain how hybrid contractarianism satisfies the condition of priority. However, I will go on to argue that the congruence between rationality and reasonableness is achieved at the cost of the condition of generality. Even if rationality and reasonableness justify the same principles, this can be done only under a substantial conception of rationality which is not generally acceptable.

6.2 Hybrid contractarianism: an elaboration

The revival of the social contract tradition is widely regarded as one of the most significant achievements of Rawls’ political philosophy.⁸ Nevertheless, his social contract has rarely been taken seriously. It is either treated as a redundant component,⁹ or taken as an incomplete theory which goes only part of the way towards achieving its aims.¹⁰ Yet the contribution of Rawls to social contract theory has long been underestimated and his contractarian model is worthy of more attention. In this section I offer an alternative interpretation of Rawls’ social contract, in order to show that his social contract is in fact a distinctive hybrid model.

and assigns this conception a foundational role. Although he does not deny that his conception of the person is revisable on the basis of coherentist reasoning, Rawls is actually a constructivist who constructs his moral theory on the basis of a particular conception of the person. See Brink (1987: 72-73). Similarly, Scanlon also observes that although Rawls has a skin of coherentism, he actually takes certain considered judgments for granted and constructs his theory on the basis of these considered judgments. See Scanlon (2003c: 155-157).

⁸ Sandel (2006: 288).

⁹ Dworkin (1973).

¹⁰ Gauthier (1986: 5) and Scanlon (1982: 110-111).

6.2.1 The problem of the orthodox interpretation

In the orthodox interpretation of Rawls' social contract, commentators place too much importance on the Original Position. The Original Position is a hypothetical choice situation that represents an 'appropriate, fair status quo' in choosing principles of justice.¹¹ Here, numerous restrictions are imposed on the hypothetical contractors, such as the *veil of ignorance*. These restrictions guarantee that irrelevant considerations are excluded when hypothetical contractors deliberate. For example, the veil of ignorance ensures that parties do not have the knowledge of all particular facts about themselves and their social and historical circumstances, including their particular conceptions of the good.¹² The purpose of this veil of ignorance is apparent. It ensures that parties in the Original Position will not be biased by knowledge of themselves and will choose fairly and impartially.¹³ '[W]hatever a person's temporal position, each is forced to choose for all'.¹⁴

Nevertheless, the disadvantage that comes with these restrictions is also obvious. Since hypothetical contractors act as the deputies of actual people, the practical reasoning of hypothetical contractors should be similar to that of actual people, so that actual people can understand why hypothetical contractors would make certain choices. However, because of the restrictions, actual people hardly think that the choices of hypothetical contractors are *their* choices. For example, behind the veil of ignorance, parties choose without knowing who they are. But in the real world, people are not similarly ignorant. Why should they agree with the choices of contractors? To them, the choice of hypothetical contractors is merely the choice that they would make if they lacked certain kinds of information. But this reasoning is

¹¹ Rawls (1999a: 11).

¹² Rawls (1999a: 118-119).

¹³ Audard (2007: 127).

¹⁴ Rawls (1999a: 121).

completely different from what they do in the real world. As Brown observes, ‘it is hard to see why an engagement that...allowed very limited information should continue to be acceptable or to be binding upon that person when he and all others like him have been greatly changed and are altogether better informed’.¹⁵

Because of the gap between hypothetical contractors and actual people, Rawls is caught in a dilemma. If he does not explain why actual people should accept the choice of hypothetical contractors, then it is unclear why actual people should take the choice of hypothetical contractors seriously. If he gives an explanation, then the Original Position becomes redundant, for, at that time, people comply with principles of justice because of this independent reason, but not because these principles would be chosen in the Original Position.¹⁶ This dilemma can explain why many other contemporary contractarians, while they adopt Rawls’ contractarian methodology, disagree with the imposition of the veil of ignorance.¹⁷ For the sake of shortening the distance between hypothetical contractors and actual people, all claim that their social contract is made in a ‘well-informed’ manner and believe that the veil of ignorance undermines the importance of the ‘contractarian element’ in the whole theory.

6.2.2 A re-interpretation of Rawlsian hypothetical contractors: free and equal persons

I suggest that Rawls’ social contract should be interpreted in an alternative way. As I showed in Chapter 3, a contractarian theory has three components: a conception of

¹⁵ Brown (1988: 444). Similar critique can also be found in Dworkin (1973) and Nagel (1973). In fact it is one of the most common critiques of Rawls’ contract theory.

¹⁶ Marshall (1975: 457-458).

¹⁷ See Gauthier (1986: 4-5), Scanlon (1982: 110-111) and Barry (1995a: 9).

practical reason, hypothetical contractors and a hypothetical agreement. The latter two elements in Rawls' theory are usually thought to be mutually disinterested parties behind the veil of ignorance and the hypothetical contract in the Original Position. I disagree with this interpretation and argue that these two elements should be seen as referring to two other ideas. Instead of parties behind the veil of ignorance, the hypothetical contractors are actually free and equal persons with two moral powers.¹⁸ Instead of the agreement in the Original Position, the hypothetical agreement is actually the public agreement of persons in the well-ordered society. The reason we should follow two principles of justice is that they represent fair terms of social cooperation among free and equal persons, not (merely) because they would be chosen by parties behind the veil of ignorance.¹⁹

This is a better reading because the dual conception of practical reason in Rawls' contractarianism is made more explicit. Though 'Rawls is not terribly explicit about his conception of reason',²⁰ his contractarianism is no doubt 'based essentially on practical reason'.²¹ Rawls recognizes that both rationality and reasonableness are dimensions of practical reason and neither can be reduced to the other.²² So neither a

¹⁸ This conception of person is sometimes thought to be a factual description of existent people. Nevertheless, it is actually an ideal conception of the person that is generally taken as 'common sense'. See Rawls (1980: 307-308).

¹⁹ The relationship between the Original Position and the view of free and equal persons can also be seen in Rawls (1993: 28). Similar interpretation is also adopted by Samuel Freeman and Joshua Cohen, though they do not emphasize the 'hybrid' nature of Rawls' social contract. See Freeman (2007a: 4) and Cohen (2003: 90).

²⁰ Darwall (1976: 170).

²¹ This is admitted by Rawls himself (Rawls 1993: 93). However, it is worth noting that Rawls seemed to change this thought in his final period. Before his final illness, Rawls planned to revise *Political Liberalism*, and one of his purposes was to change the philosophical terminologies and to delete those terminologies about practical reason. The reason that Rawls considered revising was, I guess, because he recognized that citizens may have different understandings of practical reason in liberal democratic societies. If justice as fairness presupposes a conception of practical reason, then it may no longer be a freestanding political conception. Thus Rawls wanted to bracket this question or use another way to explain normativity. However, in the end this project remained unfinished and I have no way to verify my guess. For Rawls' final revision plan, see Rawls (2005: 438-439).

²² Rawls (1999a: 496-497)

rational contract nor a reasonable contract is adequate. This is why Rawls adopts a dual conception of practical reason. Based on this conception, the characteristics of free and equal persons are defined and these characteristics can be shown in the following three aspects: their motivation, their knowledge, and their qualification.

The motivation of contractors

According to Rawls, free and equal persons have two moral powers:

- (i) One such power is the capacity for a sense of justice: it is the capacity to understand, to apply, and to act from (and not merely in accordance with) the principles of political justice that specify the fair terms of social cooperation.
- (ii) The other moral power is the capacity for a conception of the good: it is the capacity to have, to revise, and rationally to pursue a conception of the good. Such a conception is an ordered family of final ends and aims which specify a person's conception of what is of value in human life or, alternatively, of what is regarded as a fully worthwhile life.²³

These two moral powers correspond to the two aspects of practical reason: the first power corresponds to reasonableness, while the second corresponds to rationality.²⁴ They determine what motivations free and equal persons would have. Insofar as they possess the first capacity, free and equal persons have 'a strong sense of justice, an effective desire to comply with [principles of justice] and to give one another that to which they are entitled'.²⁵ Rawls sometimes says that the sense of justice exists in 'a

²³ Rawls (2001: 19). In *A Theory of Justice* Rawls also argues that these two moral powers are the distinctive features of free and equal persons in the well-ordered society (Rawls 1999a: 442).

²⁴ Rawls (1999a: 17).

²⁵ Rawls (1999a: 274).

purely formal sense'.²⁶ Nevertheless, it does not mean that the sense of justice is an empty sense of duty which can be related to any political principle. In fact, it refers specifically to the motivation to embrace principles that represent fair terms of social cooperation that would be justified to each participant.²⁷ Free and equal persons have a motivation to participate in 'terms each participant may reasonably be expected to accept, provided that everyone else likewise accepts them'.²⁸ Hence, the sense of justice can be seen as 'express[ing] a willingness...to act in relation to others on terms that they also can publicly endorse'.²⁹ Free and equal persons are willing to regulate themselves to act in a way justifiable to each other.³⁰ As Freeman explains,

[The sense of justice] means that, independent of our fellow-feeling and desires to advance human interests, we have a desire for fairness, that is, a desire to see that human interests are advanced in ways that are fair, or (what comes to the same thing in Rawls' account) in ways that are acceptable to all from a position that is reasonable and fair between them.³¹

Apart from this motivation, free and equal persons also have a motivation to realize their rational plans, which comes from their second capacity. According to Rawls, 'the rational plan for a person is the one...which he would choose with deliberative rationality'.³² Persons form plans that are made in accordance with several principles of rational choice, for example, the relative intensity of various desires having examined them carefully, and all desires arranged in a coherent plan.³³ A person's good is related to the successful realization of these rational plans. 'Our good is

²⁶ Rawls (1999a: 126).

²⁷ Rawls (1999a: 397). A similar observation is also found in Freeman, who argues that the sense of justice is, unlike the highest-order interest, a motivation to act from justice as fairness *simply* for the sake of justice. See Freeman (2003b: 282).

²⁸ Rawls (1980: 316).

²⁹ Rawls (1993: 19). Similar observation can also be found in Scanlon (2003: 165).

³⁰ Rawls (1999a: 491).

³¹ Freeman (2007c: 259).

³² Rawls (1999a: 366).

³³ Rawls (1999a: 123-124, 367-369).

determined by the plan of life that we would adopt with full deliberative rationality if the future were accurately foreseen and adequately realized in the imagination'.³⁴

Rawls further argues that free and equal persons have rational highest-order interests in realizing their two moral capacities. The formal principles of rational choice do not tell us anything about the rationality of ends: 'we cannot derive from the definition of rational plans alone *what sorts of ends* these plans are likely to encourage'.³⁵ Hence Rawls introduces the Aristotelian Principle. The Aristotelian Principle is a general psychological principle: 'other things equal, human beings enjoy the exercise of their realized capacities (their innate or trained abilities), and this enjoyment increases the more the capacity is realized, or the greater its complexity'.³⁶ The role of the Aristotelian Principle is to account for the motivations of human beings and establish that it is intrinsic to a human being's good to realize their nature. It can explain why we prefer to do some things and not others, because some things can help us train our capacities and we take more pleasure in doing some things as our capacities become more fully realised.

Moreover, Rawls offers a Kantian interpretation of human nature: persons seen as moral agents are by their nature free and equal moral persons.³⁷ Since the nature of free and equal persons is constituted by the two moral powers, these powers are more important than other capacities that we may realize, such as the aesthetic or mathematical capacity. Without aesthetic or mathematical capacities, one can still determine one's conceptions of the good, be responsible for one's actions and participate in social cooperation with other moral persons. However, people without

³⁴ Rawls (1999a: 370).

³⁵ Rawls (1999a: 372) (*Italic added by writers*).

³⁶ Rawls (1999a: 374).

³⁷ Rawls (1999a: 495).

two moral powers are not able to be free and responsible agents who are capable of mastering their wants and answering for their actions.³⁸

The Aristotelian Principle and the Kantian interpretation of human nature together explain that realizing these two moral powers is the highest-order interest of human beings. According to the Kantian interpretation, the two moral powers constitute human nature. Then, ‘from the Aristotelian Principle it follows that the expression of [human beings’] nature is a fundamental element of their good’.³⁹ Therefore, realizing and exercising the two moral powers are the ‘highest-order interests’ of human beings. Compared with other interests, these highest-order interests are ‘supremely regulative as well as effective’.⁴⁰ This implies that, ‘whenever circumstances are relevant to their fulfilment, these interests govern deliberation and conduct’.⁴¹ As some commentators point out, this motivation plays a crucial role in explaining why people should comply with principles of justice, even more than the sense of justice.⁴²

In short, free and equal people generally have two motivations, which come from their two moral capacities respectively. The first is the motivation to regulate behaviour to embrace principles that can be justified to others, which comes from the capacity for a sense of justice. The second is the motivation to realize rational life plans, which comes from the capacity for a conception of the good. Their rational

³⁸ Rawls (1975: 284).

³⁹ Rawls, *A Theory of Justice*, P. 390.

⁴⁰ Rawls (1980: 312). See also Pogge (1989: 98). Rawls changes his description of the role of the highest-order interests in *Political Liberalism*. He argues that the highest-order interests are merely *means* to protect and advance the determinate conception of the good of a person (Rawls 1993: 312-313). In this description, the highest-order interests have only instrumental value. However, in most of the earlier writings, the highest-order interests are intrinsically good. Their good is independent of the determinate conception of the good of a person. I will take the later interpretation rather than take the former here.

⁴¹ Rawls (1980: 312).

⁴² Bates (1974: 7-8).

plans of life consist of two parts: the *ends* of their plans are the highest-order interests of realizing two moral capacities, while the *means* of their plans are determined by the formal principles of rational choice. The former is concerned with behaving rightly, while the latter is concerned with pursuing the good.

The knowledge of contractors

Many people take the hypothetical contractors in Rawls' social contract to be the parties behind the veil of ignorance, and thus object that there is a huge knowledge gap between these parties and actual people. That is why both Gauthier and Scanlon assume that their hypothetical contractors are well-informed. However, this problem can be avoided if we take Rawls' social contract to mean the public agreement among free and equal persons. Free and equal persons know their particular facts which are screened out by the veil of ignorance. 'In practical affairs an individual does have the knowledge of his situation and he can, if he wishes, exploit contingencies to his advantage'.⁴³ Nevertheless, they do not do so because of the two motivations that I mentioned above. Even though he is well-informed, these motivations still 'move him to act on principles of right that would be adopted in the Original Position'.⁴⁴ Hence, in this interpretation, hypothetical contractors do not make their decision in the absence of particular information. Rather, they choose to embrace principles of justice with full information. This greatly shortens the distance between actual people and hypothetical contractors. Actual people can no longer use the knowledge gap as an excuse for failing to identify themselves with hypothetical contractors.

⁴³ Rawls (1999a: 128).

⁴⁴ Rawls (1999a: 128).

The qualification of contractors

Since free and equal persons are characterized by two moral powers, it means that any actual person who has these two moral powers is included in the contract. A person who has these two moral powers can be seen as a free and equal person who is capable of being ‘engaged in social cooperation’.⁴⁵ Having the two moral powers is ‘the sufficient condition for being entitled to equal justice’.⁴⁶ But how common are these two moral powers? There seems no doubt that nearly every person has the capacity for a conception of the good. Actual people are generally willing to plan and are capable of planning their lives rationally. Apart from this capacity, the capacity for a sense of justice is also a common feature of human beings. Many human sentiments, such as guilt, are connected with a sense of justice.⁴⁷ The sense of justice is also a presupposition of many fundamental sentiments, such as love, trust, friendship and affection.⁴⁸ Hence, the sense of justice is much more common than many people think. ‘It seems almost certain that at least the vast majority of mankind have a capacity for a sense of justice and that, for all practical purposes, one may safely assume that all men originally possess it’.⁴⁹

Given that most of the people can satisfy this minimum standard, we can infer that nearly all persons are obligated to principles of justice. In fact, it seems that this is also the ambition of the earlier Rawls.⁵⁰ He suggests that the two capacities are a part

⁴⁵ Rawls (2001: 18).

⁴⁶ Rawls (1999a: 442) and Rawls (1963: 112).

⁴⁷ Rawls (1963: 100-106).

⁴⁸ Rawls (1999a: 428).

⁴⁹ Rawls (1963: 114).

⁵⁰ This observation is also shared by Adam Swift and Stephen Mulhall, who believe that Rawls’ social contract has universal implications. See Mulhall and Swift (1996: 18-21). Nevertheless, Rawls is also quite vague at this point, since he also says in the preface that justice as fairness ‘best approximates our considered judgments of justice and constitutes the most appropriate moral basis for a democratic

of humanity and have a history that may even be given an evolutionary explanation.⁵¹ Also, he believes that principles of justice are objective since the conception of the free and equal person represents a common standpoint that *human beings* in general could endorse.

This standpoint is objective and expresses our autonomy...Thus to see our place in society from the perspective of this position is to see it *sub species aeternitatis*: it is to regard the human situation not only from all social but also from all temporal points of view. The perspective of eternity is not a perspective from a certain place beyond the world, not the point of view of a transcendent being; rather it is a certain form of thought and feeling that rational persons can adopt within the world. And having done so, they can, whatever their generation, bring together into one scheme all individual perspectives and arrive together at regulative principles that can be affirmed by everyone as he lives by them, each from his own standpoint.⁵²

Some might object that the severely handicapped who do not have the two moral powers would be excluded from the social contract.⁵³ If this is true, then Rawls' contractarianism is also vulnerable to the problem of Hobbesian contractarianism. Yet this worry is based on a misinterpretation of Rawls. Some people might not possess these two moral powers, but it is merely because they have encountered 'the consequence of unjust and impoverished social circumstances, or fortuitous contingencies'.⁵⁴ However, they still have 'a potentiality that is ordinarily realized in due course. It is this potentiality which brings the claims of justice into play'.⁵⁵ Hence, the scope of Rawls' social contract includes all people who have the *potential*

society'. Justice as fairness seems to be a specific conception of justice in democratic societies. See Rawls (1999a: xviii).

⁵¹ Rawls (1999a: 440). Similar point of also suggested by Alan Gibbard. See Gibbard (1982: 31-33).

⁵² Rawls (1999a: 514).

⁵³ Nussbaum (2006: 64-65).

⁵⁴ Rawls (1999a: 443).

⁵⁵ Rawls (1999a: 442).

to realize their two moral powers, even includes those who actually do not have these powers.⁵⁶ In short, Rawls' social contract is a social cooperative scheme among well-informed free and equal persons who have (or have the potential to realize) the two moral capacities.

6.2.3 The hypothetical contract: justice as fairness

To Rawls, a legitimate political order among these free and equal persons should be a well-ordered society, that is, 'a society in which everyone accepts and knows that the others accept the same principles of justice, and the basic social institutions satisfy and are known to satisfy these principles'.⁵⁷ Principles of justice should be able to attain public acknowledgement and to give justification to each person's reason.⁵⁸ So the next question for Rawls is: What principles of justice would be publicly agreed by these free and equal persons?

The answer is obvious. Justice as fairness is the solution for free and equal persons, from the perspectives of both rationality and reasonableness. Justice as fairness is reasonably justified because it represents an agreement which specifies fair terms of social cooperation. Free and equal persons have a sense of justice, which can be seen as a willingness to behave in a way justifiable to others. Rawls argues that this

⁵⁶ Some might further argue that although the standard is so low, some congenitally disabled people would still be excluded because they even do not have the potential to realize the two moral powers. Nevertheless, I believe the number of excluded people would be negligible because, no matter how small the chance is, it is still *possible* for congenitally disabled people to be cured and recovered in the future. Moreover, even though some people are seriously disabled to the extent that it is impossible to be recovered, some Rawlsians have recently argued that they would still be protected. Rawls's principles of justice are applied in a four-stage sequence and the claims of these disabled people would be considered in the second stage. See Starks (2007).

⁵⁷ Rawls (1999a: 397).

⁵⁸ Rawls (2007: 13-14). A similar point can also be found in Rawls (1958: 59). This emphasis on common, public knowledge can also be found in Rousseau's writings. This further proves that both Rousseau and Rawls belong to this distinctive contractarian tradition. See Cohen (2010: 40-41).

willingness is a motivation to remain in a relationship in which everyone respects others as a free and equal moral person.⁵⁹ The Original Position can elucidate what principles would be accepted among citizens who have such a relationship with one another: ‘since the Original Position situates free and equal moral persons fairly with respect to one another, any conception of justice they adopt is likewise fair’.⁶⁰ Through the Original Position, free and equal persons know that justice as fairness is the conception of justice that can protect the relationship of mutual respect. The sense of justice motivates free and equal persons to act upon the principles of justice, which are principles that could be reasonably accepted.⁶¹

Justice as fairness is also rationally justified, for it is compatible with a person’s good. Free and equal persons should choose justice as fairness because it can guarantee ‘the adequate development and full exercise of the moral powers’, which is their highest-order interest.⁶² The first principle of justice is the principle of equal basic liberties, which guarantees that people have equal political liberty, freedom of speech and assembly, liberty of conscience and the freedom of thought.⁶³ These basic liberties are crucial in the development of two moral powers. According to Rawls, political liberty and freedom of thought are necessary for developing the capacity for a sense of justice, and liberty of conscience and freedom of association are necessary for developing the capacity for a conception of the good.⁶⁴ If these basic liberties are not protected, two moral powers cannot be developed adequately. This is why free and

⁵⁹ But Rawls’ understanding is slightly different from Scanlon. Scanlon understands the relationship of mutual recognition as a relationship where everyone recognizes one another as a reason-assessing creature. Nevertheless, both of these relationships imply that one should take one another as having equal moral status with him and consider one another’s interest fairly.

⁶⁰ Rawls (1980: 310).

⁶¹ According to Rawls, being motivated by the sense of justice and being reasonable are the same thing. See Rawls (2001: 196). Virginia Held also argues that the sense of justice is a strong motivation that inclines people to be just. See Held (1976).

⁶² Rawls (2001: 104).

⁶³ Rawls (2001: 53).

⁶⁴ Rawls (2001: 45, 112-113).

equal persons should choose principles of justice which give the highest priority to these basic liberties, so that they can fully exercise their two moral powers. Therefore, 'the principles of right and justice are collectively rational; and it is in the interests of each that everyone should comply with just arrangements'.⁶⁵ A just society 'enables a good that is otherwise unattainable'.⁶⁶

Hence free and equal persons have two reasons to follow principles of justice in the well-ordered society. One is a consideration of reasonableness and another is of rationality. Both rationality and reasonableness are independent aspects of practical reason, but both of them demand that people embrace principles of justice.

The desire to act justly and the desire to express our nature as free and equal moral persons turn out to specify what is practically speaking the same desire. When someone has true beliefs and a correct understanding of the theory of justice, these two desires move him in the same way. They are both dispositions to act from precisely the same principles: namely, those that would be chosen in the original position.⁶⁷

By embracing principles of justice people can achieve the 'essential unity' of self'.⁶⁸ Rationality and reasonableness will not make incompatible demands and thus self-contradiction can be avoided. People can give unity to their lives by accomplishing this 'congruence of rationality and reasonableness'.

Now the unity is manifest in the coherence of his plan, this unity being founded on the higher-order desire to follow, in ways consistent with his sense of right and justice, the principles of

⁶⁵ Rawls (2001: 504).

⁶⁶ Pogge (1989: 100).

⁶⁷ Rawls (1999a: 501).

⁶⁸ Rawls (1999a: 493).

rational choice...in ways that justice allows, he is able to formulate and to follow a plan of life and thereby to fashion his own unity.⁶⁹

In light of why free and equal persons choose principles of justice, one should also be clear about the real function of the Original Position. The Original Position is usually regarded as a hypothetical contract that models the agreement among actual people. Yet this interpretation gives rise to a dilemma. Either the Original Position is irrelevant because the relationship between the Original Position and actual people is unclear, or it is redundant because actual people follow principles of justice for other independent reasons. However, these critiques of the Original Position are based on a misunderstanding of its function. In fact, in Rawls' contractarian model, the hypothetical contract that models the agreement among actual people is the public agreement among free and equal persons in the well-ordered society, but not the agreement made in the Original Position. In the well-ordered society, free and equal persons discuss with others and try to come to a public agreement. Since all people are free and equal, they all engage in practical reasoning when they participate in public discussion in similar ways. Despite their different ethical, religious and cultural concerns, their deliberations are framed in terms of the same frame of mind. The Original Position is a device to make this shared frame of mind explicit. The real function of the Original Position is 'a device of representation [for free and equal persons], or alternatively, a thought experiment for the purpose of public- and self-clarification'.⁷⁰ Hence the Original Position should be seen as a heuristic device that models the shared way of reasoning of contractors in the hypothetical contract, but not the hypothetical contract *itself*.

⁶⁹ Rawls (1999a: 491-492).

⁷⁰ Rawls (2001: 17).

If Rawls' contractarian theory is correctly understood, then we can understand that the Original Position is neither irrelevant nor redundant. It is relevant because the conception of a free and equal person explains the relationship between the Original Position and actual people. Since most actual people have two moral powers, they are willing to identify themselves as free and equal persons. Given that the Original Position is used for making the reasoning of free and equal persons explicit, actual people should also be able to enter into this thought experiment for the sake of clarifying the political principles that they should embrace. Thus Rawls describes the Original Position as a thought experiment that actual people can enter into at any time.

It may be helpful to observe that one or more persons can at any time enter this position, or perhaps better simulate the deliberations of this hypothetical situation, simply by reasoning in accordance with the appropriate restrictions...To say that a certain conception of justice would be chosen in the original position is equivalent to saying that rational deliberation satisfying certain conditions and restrictions would reach a certain conclusion.⁷¹

Another reason the Original Position is not redundant is because actual people have to use this device to understand what free and equal people would agree. This device represents what restrictions these free and equal persons would affirm when they deliberate about political principles. It can effectively represent and unify the characteristics of rational and reasonable deliberations in a manageable and vivid way. Through this device, principles that are both rationally and reasonably justified can be made explicit. As Rawls argues, the Original Position is a 'procedural interpretation' that clarifies practical reasoning of free and equal moral being.⁷² Without this heuristic device, actual people would only know that they were free and

⁷¹ Rawls (1999a: 119-120). A similar point can also be found in Rawls (1993: 274-275).

⁷² Rawls (1999a: 226).

equal but would not be clear about the normative implications of this conception of the person.

In conclusion, Rawls' contractarianism is a hybrid theory because it is based on a dualistic conception of practical reason. He recognizes the importance of rational interests in justification,⁷³ but he believes that reasonableness is crucial as well.⁷⁴ Thus hypothetical contractors should be both rational and reasonable in order to represent the complete practical reasoning of actual people. These two aspects of practical reason are represented respectively by the two moral capacities of free and equal persons, who are the hypothetical contractors in Rawls' theory. By showing that free and equal persons choose justice as fairness as their political principles, Rawls can show that justice as fairness is both rationally and reasonably justified. It is reasonably justified because it characterizes a relationship of mutually respect for one another as a free and equal person. It is rationally justified because acting upon it could help one advance one's highest-order interests. This dual character of hybrid contractarianism makes it distinctively different from Hobbesian and Kantian contractarianism. In the next sections, we will consider the strengths and weaknesses of this hybrid form.

⁷³ See Rawls (1999a: 347-350). It is interesting to note that, in 'Contractualism and Utilitarianism', Scanlon admits that a moral theory should deal with a motivational question, that is, whether morality is compatible with a person's good, even though the person does not intend to deal with the motivational question in this article (Scanlon 1982: 106). However, in *What We Owe to Each Other*, Scanlon rarely discusses this question. His view becomes that, once a strong moral reason is given, people are motivated and should act according to it, regardless of their interests. That is why I take Scanlon to be a Kantian contractarian and Rawls to be a hybrid contractarian.

⁷⁴ Rawls (1999a: 497-498).

6.3 Hybrid contractarianism and the condition of priority

In the previous two chapters, I showed that the problem of Hobbesian and Kantian contractarianism is that their conceptions of practical reason are one-sided. Thus they fail to satisfy the condition of priority. Unlike these two models, hybrid contractarianism relies on a more comprehensive conception of practical reason that takes both rationality and reasonableness into account. In this section, I will show that hybrid contractarianism, based on a dualistic conception of practical reason, can also satisfy the condition of priority.

6.3.1 The priority of justice as fairness

The condition of priority requires that, in a contractarian theory, all aspects of practical reason be included in the conception of practical reason. The hypothetical contractors should not capture only a part of practical reasoning. If a contractarian theory satisfies this condition, then the political principles justified by this theory can have an overriding priority over other social norms. In fact, this problem of priority is emphasized by Rawls as well, but Rawls calls it ‘the problem of stability’. In Rawls’ theory, stability has a special meaning. It does not mean social peace or a mere avoidance of conflicts.⁷⁵ Rather, it refers specifically to whether a conception of justice is both rationally and reasonably justified, whether rationality and reasonableness can be congruent with each other.⁷⁶ If principles of justice are not rationally justified, then people would always be attracted by a rational motivation to go against these principles. Also, these principles would always be vulnerable to the challenge that they severely threaten the rational plan of life of people. Hence, a

⁷⁵ Many theorists misunderstand Rawls’ idea of stability as referring to social unity and consensus. Examples are Baier (1989) and Herschovitz (2000).

⁷⁶ Rawls (1999a: 497).

conception of justice must prove its stability. People who live under this scheme would always endorse it and would not be tempted by opposing motivations.

Whether [rationality and reasonableness] are congruent is likely to be a crucial factor in determining stability...The problem is whether the regulative desire to adopt the standpoint of justice belongs to a person's own good when viewed in the light of the thin theory with no restriction on information. We should like to know that this desire is indeed rational; being rational for one, it is rational for all, and therefore no tendencies to instability exist.⁷⁷

As Paul Weithman observes, the stability of a just society is 'one of Rawls' most pressing concerns'.⁷⁸ In light of the importance of the problem of priority, we can understand why Rawls assumes that the hypothetical contractors are both rational and reasonable.

I will now explain how the political principles chosen by these hypothetical contractors can satisfy the condition of priority. There are two possibilities where a political principle fails to satisfy the condition of priority. First, people think that this political principle is unjustified because other social norms are more important, so they violate this political principle because of other social norms. Secondly, people think that this political principle is rationally (or reasonably) justified, but think that it is reasonably (or rationally) unjustified as well. Hybrid contractarianism can avoid these two cases and ensure that political principles have an overriding authority.

For the first case, people cannot use social norms as reasons for rejecting these political principles. There are various social norms in people's lives, such as religious disciplines and cultural customs. As I discussed in Chapter 2, rationality

⁷⁷ Rawls (1999a: 497).

⁷⁸ Weithman (2010: 5).

and reasonableness do not have the same status as these social norms. Rationality and reasonableness govern the *a priori* structure of our practical reasoning and guide our thoughts, whereas social norms are only elements that we consider during practical reasoning. When people engage in practical reasoning and think about what political principles they should choose, they consider these social norms because these are things that occupy crucial positions in their lives. Under the guidance of rationality and reasonableness, they look for a balance between different norms as they also evaluate these norms. Eventually their practical reasoning arrives at the conclusion whereby their society should be governed by principles of justice which have a priority over other social norms. As long as this conclusion is reached, they are unjustified in rejecting this order of priority by referring to the importance of other social norms, since the importance of these social norms has already been taken into account in practical reasoning and has been outweighed by other considerations. Thus these norms no longer constitute a reason for overthrowing the conclusion. If people refuse to acknowledge the priority of principles of justice, then they are simply ignoring the fact that no reason can justify their action.

For example, suppose that, in a society governed by principles of justice, the places of secondary school are fairly allocated by the government (principles of fair equality of opportunity). Now there is a man who is the government officer responsible for this allocation and he has a chance to secretly privilege his child to enter into a high-ranking secondary school. As a parent, he has an obligation to give his child the best possible start. But is this familial obligation a reason for action? From the reasonable perspective, privileging his child is unjustified since other candidate students would be unfairly situated. From the rational perspective, it is unjustified as well because the benefits gained from doing this is outweighed by the good constituted by the highest-order interest in developing the capacity for a sense of justice. In both of these two aspects of practical reason, the familial obligation is superseded by some other considerations. Given that a consideration can be justified as a reason for action only when it is either rationally justified or reasonably justified, the familial

obligation fails to count as a reason for action, even though it has a great weight in people's lives.

Hybrid contractarianism can avoid the second case too because political principles chosen by the hypothetical contractors are both rationally and reasonably justified; justice represents the congruence of rationality and reasonableness. Its strength can be made explicit by comparing the hybrid contractarian model with Hobbesian and Kantian contractarianism, both of which fail to avoid the second case. Hobbesian and Kantian contractarianism are flawed because their conceptions of practical reason are *incomplete*; Hobbesian contractarianism justifies political principles merely by rationality, whereas Kantian contractarianism justifies political principles merely by reasonableness. The choice of their hypothetical contractors overlooks reasons justified by another aspect of practical reason. For actual people, they have a reason to follow the political principles justified by these two models, but they have a reason to reject these political principles as well. Insofar as neither reason is conclusive, either embracing these political principles or rejecting them is justifiable.

Unlike Hobbesian and Kantian contractarianism, hybrid contractarianism has a more *complete* conception of practical reason and discusses both rational and reasonable deliberation. From the reasonable perspective, principles of justice are justified because they respect each person as a free and equal being and govern in a way which can be publicly justified. From the rational perspective, principles of justice are justified because following them would be the most effective way to realize people's two highest-order interests. Rationality and reasonableness are congruent with each other and both of them justify the loyalty to principles of justice. Insofar as there are no other aspects of practical reason except rationality and reasonableness, principles of justice represent the outcome of the complete process of practical reasoning. So actual people no longer have reason to reject these principles because all reasons are justified by either rationality or reasonableness, and these two aspects

of practical reason both demand that they should embrace these principles. Hence, given that social norms cannot challenge principles of justice and principles of justice represent a common requirement of rationality and reasonableness, these principles can achieve an overriding political authority.

6.3.2 The conflict between two highest-order interests

However, I believe that Rawls is too optimistic that the two highest-order interests always lead to the same disposition. According to Rawls, free and equal persons have two highest-ordered interests, one is to realize the capacity for a conception of the good, and the other is to realize the capacity for a sense of justice. Even if we assume that these two interests are really the highest-order interests, they may still conflict with each other, and Rawls seems to overlook this possibility. I will focus on the capacity for a conception of the good and argue that embracing principles of justice may not be the best way to develop this capacity.

In Rawls' definition, the capacity for a conception of the good is not only a capacity to pursue a conception of the good, but also a capacity to *revise* a conception of the good, in order to ensure that one's current conception of the good is the one most preferred.⁷⁹ People who are interested in realizing this capacity do not view themselves as inevitably tied to the pursuit of any particular conception of the good. In order to realize this capacity, free and equal persons choose to act upon principles of justice that guarantee an adequate quantity of primary goods. The difference principle provides an ideal circumstance for people to pursue the interest of realizing the capacity for a conception of the good, because it guarantees that even the least advantaged in society can still possess an adequate quantity of primary goods. These

⁷⁹ Rawls (2001: 19). Similar point is also emphasized by Buchanan (1975).

primary goods would allow one to have enough all-purpose means to choose a conception of the good from a variety of possibilities.

Nevertheless, if one really aims at pursuing this highest-order interest, then one should not be satisfied with the distribution of primary goods which is determined by the difference principle. From the definition of the capacity for a conception of the good, we can say that this capacity is more realized if people are able to revise their conception of the good more thoroughly. When people revise their conception of the good, they cannot look only at their current conceptions. On the contrary, they need to compare their current conceptions with others, so that they can know whether the current conceptions of the good are the ideals that they most prefer. This means that if the choices of alternative conceptions of the good are limited, then the revision of their conceptions is also limited, for their current conception is not open to too many challenges by other conceptions of the good. Hence, the more conceptions of the good people can choose and compare, the deeper these people can revise their conceptions.

On the other hand, if one has more primary goods, more conceptions of the good are available for one to choose. Some conceptions of the good, such as musician, artist or traveller, require plenty of primary goods to realize. These conceptions of the good are impossible for a person to realize if they do not have enough primary goods. Moreover, if one has more primary goods, then one is better able to acquire more knowledge. As long as one can gain more knowledge, one has more resources to reflect upon one's conception of the good. People can hardly say that they have reflected upon their conceptions of the good thoroughly if they lack a rich knowledge of their own conceptions, and other conceptions. This rich knowledge depends on abundant primary goods. All these facts imply that the more primary goods one has, the more conceptions of the good from which one is able to choose.

Now we can see that there is a relationship between the quantity of primary goods and the freedom of a person. One is freer if one can reflect upon one's conception of the good more thoroughly, because people are free if they are capable of standing apart from their conceptions and reflecting, and are freer if this capacity is more developed. This capacity is more developed and the conception of the good of a person can be revised more deeply if that person has more conceptions of the good to choose from and to compare. Therefore, the value of developing this moral capacity can also be explained by freedom. The more primary goods one has, the further the capacity for a conception of the good one can realize, the freer one is.

If this is true, the interest of realizing the capacity for a conception of the good may conflict with the interest of realizing the capacity for a sense of justice. The capacity for a sense of justice is 'the capacity to understand, to apply, and to act from...the principles of justice that specify the fair terms of social cooperation'.⁸⁰ Since this capacity can be realized only by living in a well-ordered society, the interest in developing the capacity for a sense of justice motivates people to comply with principles of justice unconditionally.⁸¹ Unlike this capacity, the development of the capacity for a conception of the good depends on how many primary goods one possesses. Suppose one could get more primary goods by violating the difference principle, then one should go ahead for the sake of realizing one's capacity for a conception of the good further. So the interest in realizing the capacity for a conception of the good can become a motivation to violate principles of justice in certain cases. According to Rawls, these two highest-order interests have the same status.⁸² Since both of these two interests are the highest-order interests, pursuing either of them are rational to people. Some rational people may choose to realize the capacity for a conception of the good but not the capacity for a sense of justice.

⁸⁰ Rawls (2001: 19).

⁸¹ Rawls (1999a: 404).

⁸² Rawls (1980: 312).

Hence, on certain occasions, although it is reasonable for people to comply with principles of justice, it is rational for people to violate them for the sake of getting more primary goods. Rationality and reasonableness fail to be congruent with each other.

6.3.3 The maximizing nature of the highest-order interest

Perhaps Rawls might respond that the interest in realizing the capacity for a conception of the good cannot be ‘maximized’. In fact, in *Political Liberalism*, Rawls denies that the pursuit of these interests is a kind of maximization. The two moral powers cannot be maximized. ‘A coherent notion of what is to be maximized is lacking. People cannot maximize the development and exercise of two moral powers at once’.⁸³ Even if maximizing the development of these two moral powers is possible, people have no intention to maximize. Rather, they are interested only in realizing the two moral powers to an adequate extent ‘as long as their powers and abilities suffice for them to be normal cooperating members of society’.⁸⁴ Once this condition is met, they have no interest in developing the two moral powers. Therefore, their highest-order interests would be satisfied if their two moral powers were *adequately* developed.

Rawls’ description may be correct for the capacity for a sense of justice. It is impossible to maximize the development of this capacity since this ability cannot be improved anymore after one possesses it. All people can be separated into two kinds: those who have a sense of justice, and those who do not. However, I do not see why the development of the capacity for a conception of the good cannot be maximized.

⁸³ Rawls (1993: 333).

⁸⁴ Rawls (2001: 60).

The capacity for a conception of the good is clearly different from the capacity for a sense of justice, because this can be developed further after one possesses it. Different people can have different degrees of maturity in their capacity for a conception of the good. People's capacity for a conception of the good is more mature if they can reflect upon their conception of the good more thoroughly. For example, if they have more freedom, wealth and knowledge, then they can compare their conceptions of the good with differing conceptions across diverse areas.

Given that the pursuit of the highest-order interests is a fundamental good of human beings, one should have an intention to maximize the development of the capacity for a conception of the good, and not be satisfied with merely an adequate level of development. According to the definition of highest-order interests, they are 'supremely regulative as well as effective'.⁸⁵ If these highest-order interests are really so important, then why should people be satisfied and stop pursuing them after pursuing them to a certain degree? This can also explain why parties in the Original Position adopt an instrumental conception of rationality, preferring 'more primary social goods rather than less'.⁸⁶ For more primary goods can secure further development of the capacity for a conception of the good.

Rawls might respond that they would be satisfied with the adequate development of the two moral powers because developing these capacities is merely for the sake of taking part in social cooperation.⁸⁷ These capacities are means to advance their conceptions of the good only. But if this is true, then acting upon principles of justice becomes a 'necessary evil' to join the social cooperative scheme and realize one's own conception of the good, but not an intrinsic good for a person. This contradicts

⁸⁵ Rawls (1980: 312).

⁸⁶ Rawls (1999a: 123).

⁸⁷ Rawls briefly discusses this point in *Political Liberalism* (Rawls 1993: 333-334).

what Rawls says about the rational good of being just. According to Rawls, the realization of the two moral capacities is intrinsically good insofar as ‘it is rational for persons to want in themselves and in one another as things appreciated for their own sake’.⁸⁸ That is why free and equal persons are willing to revise their conceptions of the good if these conceptions are incompatible with the two highest-order interests. If the two moral powers are merely a means to take part in social cooperation, then the most rational strategy of people should be pretending to be people who possess the two moral powers, keeping their conception of the good unchanged and waiting for chances to break the rules. This is no doubt not the strategy proposed by Rawls.

Nevertheless, this problem can be avoided as long as Rawls adds an assumption of priority between the two interests. Rawls could say that human nature is constituted only by the capacity for a sense of justice. In this new interpretation, people take the capacity for a sense of justice as more important than the capacity for a conception of the good. Even though reflection on the conception of the good is crucial, it can only be done within limits and persons can only revise their conceptions of the good by a fair share of primary goods in their hands. Suppose that this priority is justified, the interest in realizing the capacity for a sense of justice would become the only highest-order interest, while the interest in realizing the capacity for a conception of the good would become at most the second highest-order interest. Now the rational disposition of individuals would be to comply with the principles of justice in order to realize their sense of justice, which would not be incompatible with principles of justice that could be reasonably justified. This additional assumption can rescue the congruence between rationality and reasonableness and the condition of priority can be satisfied.

⁸⁸ Rawls (1999a: 463). Similar observation is also found in Nagel (2003: 83).

In this section I discussed whether or not Rawls' ambition to fulfil the condition of priority is successful. Rawls aims at satisfying the condition of priority by showing that both rationality and reasonableness justify principles of justice. Provided that all reasons are justified by either rationality or reasonableness, people have no reason for rejecting principles that are both rationally and reasonably justified. But Rawls overlooks the fact that the interest in realizing the capacity for a conception of the good, which is one of the rational highest-order interest, may conflict with principles of justice. This conflict harms the congruence of rationality and reasonableness. However, I also argue that this problem can be solved by adding a further assumption of priority between two highest-order interests. By assuming that the interest in realizing the capacity for a sense of justice can outweigh the interest in realizing the capacity for a conception of the good, the rational choice of people can be guaranteed to be compatible with reasonableness. Thus, we can see that Rawls' ambition can be satisfied, provided an extra assumption is added. In the next section, however, we will see the cost of this ambition is that Rawls' theory fails to fulfil the condition of generality, and the extra assumption would make his theory more controversial.

6.4 The cost of satisfying the condition of priority

Hybrid contractarianism is better than Hobbesian and Kantian contractarianism since it can fulfil the condition of priority that the other two kinds of contractarianism fail to satisfy. However, the condition of priority is not satisfied without any cost. The congruence between rationality and reasonableness is based on a substantial conception of rationality. Since this substantial conception of rationality is a part of the conception of practical reason of Rawls' theory, this undermines the possibility of Rawls' theory satisfying the condition of generality since this conception of rationality is not a weak assumption that is generally acceptable. In this section, we can see that although the substantial interpretation is the key to achieve the condition of priority, it is also the cause of failing to achieve the condition of generality.

6.4.1 The problem of generality

The strength of Hobbesian contractarianism is that it is based on a weak, uncontroversial conception of rationality. This strength allows it to satisfy the condition of generality, yet it also renders it unable to satisfy the condition of priority. In order to satisfy the condition of priority and achieve the congruence between rationality and reasonableness, Rawls adopts a more substantial conception of rationality. This raises the question of whether this conception can satisfy the condition of generality, that is, it may not be generally accepted by actual people who are bound by principles of justice.⁸⁹ His conception of rationality consists of three components: the formal principles of rational choice, the Aristotelian Principle and the Kantian interpretation of human nature. For the first component, we have seen in Chapter 4 that the formal principles of rational choice are generally accepted as guiding principles of practical reason. The second component is also not a problem. Rawls is confident that the Aristotelian Principle is psychologically self-evident: ‘We need not explain here why the Aristotelian Principle is true’.⁹⁰

But someone might cast doubt on the acceptability of the Kantian interpretation. Rawls supposes that, if people were brought up in a well-ordered society, then they would identify themselves as free and equal persons who would accept the Kantian interpretation. They would take the two moral capacities as the two most

⁸⁹ However, the conception of reasonableness can satisfy the condition of generality, for Rawls has already explained how the sense of justice could have emerged when people were brought up in a well-ordered society. Through a three stages procedure (morality of authority, morality of association and morality of principle), people can be developed into a reasonable person who has an effective sense of justice (Rawls 1999a: 405-419). Also, Rawls’ conception of reasonableness is similar to Scanlon’s, and in Chapter 5 we already saw that Scanlon’s conception of reasonableness satisfies the condition of generality. So we can assume that Rawls’ conception of reasonableness can fulfill the condition of generality as well.

⁹⁰ Rawls (1999a: 374).

fundamental capacities that constitute their nature.⁹¹ However, why is it rational for people to accept this understanding of human nature when there are countless interpretations of human nature? Even if actual people live in a well-ordered society, they may still develop highly different understandings of human nature. Some people may adopt a Christian interpretation of human nature, and take the spiritual capacity as the most important. Some people may adopt a romantic interpretation of human nature, and take the artistic and imaginative capacity as the most important. There is no single rational answer to this question. People have no specific reasons for adopting the Kantian interpretation of human nature. If this is so, then realizing the two moral capacities is not their highest-order interests as well. They may be more interested in realizing other capacities and thus their most rational choice is not to embrace principles of justice. Hence, it seems too optimistic to take it for granted that the Kantian interpretation would be widely accepted by actual people after rational deliberation.

Moreover, according to my discussion in the last section, we can see that if the condition of priority has to be satisfied, then Rawls has to add a further assumption of priority between two highest-order interests in the Kantian interpretation. This additional assumption would further undermine the acceptability of the conception of rationality. In this new interpretation of human nature, it is rational for people to take the capacity for a sense of justice as their most fundamental capacity that constitutes their nature; hence the highest-order interest is to realize this capacity. However, this new interpretation is more unacceptable than the original Kantian interpretation of human nature. Although it is also controversial, the Kantian interpretation is at the least appealing because it emphasizes freedom as a part of human nature. As Rawls notes, the capacity for a conception of the good shows the freedom of a person,

⁹¹ Rawls argues that this belief of human nature should be generally shared in a well-ordered society. See Rawls (1980: 324).

because this person ‘views [himself] as independent from and not identified with any particular conception of the good, or scheme of final ends’.⁹² He is not determined by these contingent elements but be able to ‘free’ himself from them to determine what he is. However, in the new interpretation, this capacity is excluded from human nature. The new interpretation takes the capacity for a sense of justice to be the most fundamental capacity. Hence people are by nature ‘just’ or ‘reasonable’, but not ‘free’. This goes against the ordinary thinking of many people, since many people usually think that freedom is the distinctive characteristic of human beings. Therefore, although assuming the capacity for a sense of justice as the most fundamental capacity can secure the congruence between rationality and reasonableness, the cost of making this assumption is that the conception of rationality would become less appealing. The cost of satisfying the condition of priority is much larger than Rawls expects.

From this we can see the congruence between rationality and reasonableness presupposes a narrow definition of rationality. People not only have to agree on the Kantian interpretation of human nature, but also on the priority between two highest-order interests. Given that in a well-ordered society citizens enjoy freedom of thought and conscience, citizens would develop their own understanding of human nature. A conception of rationality that relies on a Kantian interpretation of human nature can hardly prevail in a well-ordered society. To those people who do not share Rawls’ conception of rationality, the rational choice of the hypothetical contractors in hybrid contractarianism is strange to them. They would fail to identify themselves with hypothetical contractors and thus would not think principles of justice have normative force. That is why Rawls no longer holds this substantial conception of

⁹² Rawls (2001: 21).

rationality in his later writings, since he also recognizes that this conception undermines the acceptability of his social contract.⁹³

6.4.2 Defence (1): Kantian interpretation as a fact

The negative effect of abandoning this substantial conception of rationality will be discussed at the end of this chapter. However, before I move on to discuss how Rawls modifies his social contract in his later writings, I believe that the earlier version of his social contract needs fuller examination. I argued earlier that Rawls' conception of rationality is controversial because it presupposes a specific interpretation of human nature. To this problem Rawls has two possible responses. First, he could state that human nature is a fact, so whether people reach a consensus on it does not matter. Secondly, Rawls could argue that the two moral capacities are the necessary condition for pursuing any worthwhile way of life, so that all rational persons should take the two moral capacities as more fundamental human capacities. I will examine these two possible responses in this and the next sections.

First, Rawls might simply bite the bullet and argue that the assumption of human nature is not a belief that people can choose not to hold, but rather a fact that awaits discovery. People are by nature free and equal not because they should understand themselves as such a being, but rather because they are *in fact* such a being. What they need to do is merely recognize this fact and take realizing this nature as their highest-order interest. Actually, this seems to be Rawls' attitude in his earlier writings.⁹⁴ He does not bother too much with the question of why people should accept this interpretation, but simply states that this interpretation is one of the

⁹³ Rawls (1993: xvi).

⁹⁴ Similar observations can also be found in Pogge (1989: 98-99).

‘premises [that] are the elementary facts about person’.⁹⁵ As he says, the self as a free and equal person is like our ‘noumenal self’.

The description of the original position resembles the point of view of noumenal selves, of what it means to be a free and equal rational being. Our nature as such beings is displayed when we act from the principles we would choose when this nature is reflected in the conditions determining the choice.⁹⁶

Hence, a supporter of the earlier Rawls might argue that the later Rawls gives up the Kantian interpretation too quickly. The later Rawls wrongly thinks that human nature is something that is chosen. Actually, this is a general fact and we can only acknowledge it.

However, claiming that certain interpretations of human nature as a fact simply avoids the question, for no matter whether or not this interpretation is a fact, it is still generally unacceptable to many people. The problem of generality is not whether this interpretation is true, but why actual people should take it as true. If actual people do not share this interpretation, then they will have an alternative conception of rationality. They will not think that justice as fairness is rationally justified to them. So they will still question the normativity of hypothetical contract. The problem of failing to fulfil the condition of generality still exists.

Moreover, this response would make Rawls’ social contract lose an important virtue of contractarianism. One of the virtues of contractarianism is that it emphasizes a respect for the individual’s will. A contractarian political principle should be based

⁹⁵ Rawls (1999a: 226). Similar point is also reflected in Rawls’ interpretation of Kant, that Kant simply takes the conception of free and equal person for granted and does not justify it. See Rawls (1989a: 513-514).

⁹⁶ Rawls (1999a: 225).

on a premise that is mutually recognized.⁹⁷ However, in this response, a certain interpretation of human nature is presupposed to be true, regardless of what actual people really think. Even if actual people do not appreciate the rational goods assumed by Rawls after the most careful reflection and deliberation, they will still be bound by principles of justice. The rational justifiability of principles of justice will be unaffected by the strong disagreement of many people. In this case, it seems hard to say that principles of justice take the will of people seriously. Taking a certain interpretation of human nature as objectively true would serve only to cast an authoritarian shadow on Rawls' contractarianism.

6.4.3 Defence (2): the necessity of two moral capacities

The failure of the first response shows that Rawls must give a reason to justify why people should accept his interpretation of human nature. This leads to the second possible response to my critique. Although Rawls does not explain why people should accept the Kantian interpretation of human nature explicitly, his student, Samuel Freeman, tries to take up the task. According to Freeman, these two moral capacities are the necessary conditions for people to pursue any worthwhile way of life.⁹⁸ Since humans are social beings, they can only pursue their way of life in the context of social cooperation. But those who are capable of taking part in social cooperation must have two moral capacities, so that they are 'free and responsible agents capable of controlling their wants and answering for their actions'.⁹⁹ The

⁹⁷ This point is also emphasized by Rawls himself. See Rawls (1999a: 12, 508).

⁹⁸ In fact, a similar point is also mentioned by Rawls in 'Kantian Constructivism in Moral Theory', even though he is not talking about the question of human nature. Rawls argues that 'Reasonable presupposes Rational', for a rational conception of the good can only be realized within a society. So rational conceptions of the good can only be realized within the limit of principles of justice: 'Reasonableness subordinates Rational'. However, Rawls discusses this point only briefly. See Rawls (1980: 317) and also Rawls (1993: 312-313).

⁹⁹ Freeman (2003: 296)

capacity for a conception of the good guarantees that people are capable of being responsible for the projects that *belong* to them.¹⁰⁰ The capacity for a sense of justice guarantees that people are capable of respecting one another and taking part in fair terms of social cooperation. These two moral powers are the necessary conditions for being a member of society. ‘A person without these capacities is not recognized by others as answerable for his or her acts or ends (morally or legally) or deemed capable of taking an active part in social cooperation’.¹⁰¹

[D]evelopment of the sense of justice (along with the capacity for a conception of the good) is a condition of individuals being rational moral agents who are capable of assuming responsibility for their actions and taking part in, and benefiting from, social life. People who do not develop their capacities for music or sports, while they may miss out on worthwhile activities, can nonetheless lead good lives engaged in other pursuits. But those whose moral capacities for justice (and the capacity to be rational) remain undeveloped are not capable of social life. They are not then in a position to achieve the benefits of society and will be hard pressed to learn and pursue most any worthwhile way of life.¹⁰²

In light of Freeman’s argument, we can see why it is rational for people to take the two moral capacities as constituting their nature. It is true that, after rational deliberation, people may develop different self-understandings. However, no matter how they understand themselves and what they believe, they still have to live within a society. From the practical perspective, they have to be free and equal persons who can participate in social cooperation first, and then becoming the kind of persons they want to be. Although Freeman’s argument still fails to solve the conflict between two highest-order interests, at least it shows that it is always rational to take realizing the two moral capacities as the highest-order interests.

¹⁰⁰ Rawls (2001: 21-22).

¹⁰¹ Freeman (2001: 295). A similar point can also be found in Rawls (1993: 73-74).

¹⁰² Freeman (2003: 296).

However, even if we ignore the conflict between two highest-order interests, Freeman's argument can at most prove that people have an *instrumental* reason to develop the two moral capacities; yet they do not have an *intrinsic* reason to do so. Realizing the two moral powers is the precondition of participating in social cooperation. But this means that if people want to enjoy the benefits of society, they should realize these two moral capacities. People can agree that they have to realize the two moral capacities, but still refuse to treat these capacities as central to their 'nature'. They can take other capacities, such as the aesthetic capacity, as the most important capacity that constitutes their nature, and then take the two moral capacities as their means to realize the aesthetic capacity. So, according to the Aristotelian Principle, their highest-order interest would come when realizing the aesthetic capacity. The two moral capacities are only effective means for them to participate in social cooperation in order to develop their aesthetic capacity.

This also means that, if there is a chance that people could violate the principles of justice without being known by others, then they have no reason not to do so. To them, following principles of justice is necessary because their aesthetic capacity could be fully realized in a context of social cooperation. However, if they violated the principles of justice, they could get more primary goods to realize their aesthetic capacity. But at the same time, they could also stay in the well-ordered society and enjoy the benefits of that society. From a rational perspective, they should break the rules and free-ride off of others' efforts as long as this would go unnoticed, even though this implies that the realization of their two moral capacities would be harmed. We can see the flaw in Freeman's argument here. Freeman only proves that the two moral powers are the necessary condition for participating in social cooperation. Yet it does not necessarily mean that the two moral powers constitute the 'nature' of people. It is possible that people think very differently: not that the value of other capacities depends on the two moral capacities, but rather, the value of the two moral capacities depends on other capacities.

The failure of Freeman's argument sheds light on the fundamental problem of the Kantian interpretation. The problem with this interpretation is not that it is a mistaken understanding of human nature, but rather that this understanding itself is something beyond rationality. Given that actual people have freedom of thought and conscience in the well-ordered society, they can have different views on human nature after rational deliberation. These views on human nature will then determine different attitudes towards the two moral capacities. This implies that there will be no unified views on the question of human nature and highest-order interests. Through understanding the controversy over the question of human nature, we can understand why Rawls gives up on the Kantian interpretation in his later writings, since assuming the Kantian interpretation implies assuming an impossible consensus in the well-ordered society.

Therefore, although Rawls' hybrid contractarianism can satisfy the condition of priority, this is done at the cost of the condition of generality. In the last section, I showed that the congruence between rationality and reasonableness must presuppose a Kantian interpretation of human nature and an additional assumption of priority between two highest-order interests. The purpose of this section is to examine whether or not the conception of practical reason which includes these two presuppositions is generally acceptable. Since the truth of the Kantian interpretation of human nature is something that cannot be rationally justified, this interpretation is too controversial, and many actual people may refuse to adopt it in the well-ordered society. The conception of practical reason which presupposes the Kantian interpretation fails to fulfil the condition of generality, let alone the assumption that further presupposes the assumption of priority between two highest-order interests. When too many controversial assumptions are built into the conception of practical reason, this greatly undermines the possibility that actual people would identify with the hypothetical contractors.

6.5 Conclusion

The purpose of this chapter is two-fold. The first purpose is to show the first attempt of Rawls to construct a hybrid contractarian model. In the previous two chapters, I showed that both Hobbesian and Kantian contractarianism fail to satisfy the condition of priority because they overlook one of the aspects of practical reason. Only hybrid contractarianism can avoid their weakness because it adopts a dualistic conception of practical reason. By showing the congruence of rationality and reasonableness, the condition of priority can be satisfied. Nevertheless, the congruence of rationality and reasonableness also creates a new problem for hybrid contractarianism, and considering this is the second purpose of this chapter. The congruence relies heavily on a substantial conception of rationality. This conception is more substantial than Rawls himself expects, for the congruence is possible only when a further assumption of priority is added. However, if the conception of rationality is so substantial, then it will also be highly controversial. Hobbesian contractarianism can satisfy the condition of generality because its formal conception of rationality is generally acceptable. But Rawls' conception of rationality is more substantial, since it presupposes the Kantian interpretation of human nature. Hence it faces a bigger difficulty in fulfilling the condition of generality.

In fact, Rawls also acknowledge this problem in *Political Liberalism*. A well-ordered society is characterized by 'reasonable pluralism'.¹⁰³ In this society, insofar as people have freedom of thought and conscience, it is unrealistic to expect that they will all agree in their religious, philosophical or ethical beliefs.¹⁰⁴ Given that the Kantian interpretation of human nature is only one of these ethical beliefs, actual people may hold, alternative beliefs and their highest-order interests may not express their nature

¹⁰³ Rawls (1993: 3-4).

¹⁰⁴ Rawls (1993: 54-55).

as free, equal and moral beings. So it is ‘unrealistic’ to expect that the Kantian interpretation would be widely accepted by actual people.¹⁰⁵ This is why Rawls no longer appeals to this interpretation when he talks about the rational good of justice as fairness in his later writings.¹⁰⁶

However, dropping the Kantian interpretation would create other negative effects. Without the Kantian interpretation of human nature, Rawls can depend only on the Aristotelian Principle, arguing that justice is *a good* because it helps us to realize our two moral powers and that human beings can always gain enjoyment from realized capacities.¹⁰⁷ But now the two moral powers are only *two* of the *many* human capacities. They are not necessarily more important than other human capacities. So why should people choose the good of justice but not the other goods which they could enjoy if they realized other human capacities? Rawls can at most argue that acting upon principles of justice can satisfy particular interests, but fails to show why these particular interests are ‘highest-ordered interests’. The condition of priority fails again, for now rationality and reasonableness leads to a different conclusion. Rawls is caught in a dilemma here: if the conception of rationality incorporates the Kantian interpretation of human nature, then Rawls’ contractarianism fails to satisfy the condition of generality; but if the conception of rationality excludes the Kantian interpretations of human nature, then Rawls’ contractarianism fails to satisfy the condition of priority.

Rawls’ strategy is to find the third way, and it is also one of his main tasks in his later writings. In *Political Liberalism*, he abandons the Kantian interpretation of human nature, though he maintains that principles of justice have priority over other

¹⁰⁵ Rawls (1993: xvi).

¹⁰⁶ Rawls (2001: 200).

¹⁰⁷ Rawls (2001: 200-201).

considerations. But now the priority argument relies on a freestanding political conception of free and equal citizens. Rawls turns to argue that, in a more restricted ‘political’ realm, reasonableness can be congruent with rationality, and these two conceptions of practical reason are still generally accepted. In the next chapter, we will consider whether or not the strategy of the later Rawls is successful.

Chapter 7 Hybrid Contractarianism (2): the later Rawls

7.1 Introduction

As a hybrid contractarian, Rawls adopts a dual conception of practical reason and argues that principles of justice as fairness are justified from both rational and reasonable perspectives. Embracing principles of justice represents a way of treating others reasonably, as well as a way of realizing one's highest-order interests rationally. Through embracing principles of justice, practical reason can be unified, since rationality can be congruent with reasonableness. However, this congruence between rationality and reasonableness presupposes a substantial conception of rationality, which assumes that the realization of two moral powers is the highest-order interest of everyone. This substantial conception of rationality is highly controversial and seriously threatens the general acceptability of hybrid contractarianism even in a well-ordered society.

Rawls recognizes this problem and acknowledges in *Political Liberalism* that his original account of the congruence between rationality and reasonableness 'is not consistent with [A *Theory of Justice*] as a whole'.¹ The reason is that, given that freedom of thought and conscience are guaranteed, reasonable pluralism would occur.² Citizens would develop 'a diversity of opposing and irreconcilable religious, philosophical and moral doctrines'.³ It is a natural result of the exercise of human reason within the framework of the liberal democratic society. 'No one of these doctrines is affirmed by citizens generally'.⁴ Hence, the Kantian comprehensive

¹ Rawls (1993: xviii).

² Rawls (1993: 36).

³ Rawls (1993: 3-4).

⁴ Rawls (1993: xviii) and Rawls (2001: 187).

doctrine, which is only one of these many comprehensive doctrines, is unlikely to be generally accepted unless the government uses oppressive measures.⁵ Rawls' conception of rationality, which relies heavily on the Kantian interpretation of human nature, must be revised.

However, Rawls does not give up the ambition of proving the congruence between rationality and reasonableness, and now takes it even more seriously.⁶ His later writings are mainly concerned with this question: How to maintain the congruence between rationality and reasonableness without relying on a controversial conception of practical reason? The strategy of the later Rawls is to define several 'political conceptions', which represent a specific context of applying principles of rationality and reasonableness.⁷ Rationality and reasonableness can be congruent with each other in this particular context. He no longer assumes that people are by nature free and equal and principles of justice are the most effective way to realize human nature. Rather, he argues that these political conceptions, which are the ground of the congruence argument, are only freestanding conceptions which are compatible with a wide range of reasonable comprehensive doctrines. Thus principles of justice, which are justified by these political conceptions, become the focus of an overlapping consensus in a reasonably pluralistic society. Hence, instead of arguing that justice as fairness represents a fundamental human good, the later Rawls rather chooses to emphasize the 'freestanding' character of political conceptions.

The purpose of this chapter is to discuss whether or not this new strategy can rescue Rawls' hybrid contractarianism. First I will briefly discuss the changes made by the later Rawls. Then I will show that, although the later Rawls gives up the Kantian

⁵ Rawls (1993: 37).

⁶ Klosko (1994: 1886-1888).

⁷ Rawls (1993: 107).

interpretation, the political conceptions are still highly controversial even in a well-ordered society. Although Rawls believes that citizens in the well-ordered society would generally accept political conceptions, I argue that, with the free exercise of human reason, citizens would also be likely to question the idea that political conceptions should be the *sole* basis of the justification of political principles. After that I will examine several possible defences from Rawls and conclude by showing that the later Rawls still cannot avoid a dilemma between generality and priority. The scope of his contract is actually much more limited than Rawls himself anticipates. Although Rawls' hybrid contractarianism can satisfy the condition of priority, the cost of doing so is the failure to fulfil the condition of generality.

7.2 The revision of hybrid contractarianism: overlapping consensus

It is commonly accepted that Rawls makes a 'political turn' in his later writings. He tries to get rid of the Kantian comprehensive doctrine presupposed in his earlier writings and emphasizes that his contractarian theory is in fact based on political conceptions that are widely shared among members of the liberal democratic societies. Although members have different reasonable comprehensive doctrines, they still share certain political conceptions which are used for the common ground of public justification. The stability of a well-ordered society is guaranteed by this 'overlapping consensus'. In this section, I will discuss how this 'political turn' affects the contractarian theory of the later Rawls.

7.2.1 The change of hypothetical contractors: scope and motivation

In general, the contractarian theory of the later Rawls has changed in two aspects; one concerning the scope of contract and the other concerning the motivation of contractors. First, he acknowledges that justice as fairness is a set of political

principles only for liberal democratic societies. Secondly, the motivations of contractors are now based on some political conceptions, particularly the political conception of free and equal citizenship, which is widely shared among members of liberal democratic societies. I will illustrate these two changes in turn in this section.

The scope of contractors

In *A Theory of Justice*, Rawls is ambitious in proposing that principles of justice can hold for all human beings. Since two moral capacities are the general characteristics of human beings, principles of justice should be able to be applied in all societies. However, in *Political Liberalism*, Rawls gives up this ambition and argues that principles of justice can only be justifiably enforced in liberal democratic societies.⁸ It does not mean that Rawls no longer thinks that justice as fairness is a universal value and every society in the world ought to strive to become a liberal democratic society.⁹ However, he acknowledges that justice as fairness presupposes a public political culture which does not exist in non-liberal-democratic societies.

As we saw in the last chapter, justice as fairness is based on a conception of free and equal persons who possess two moral capacities. Rawls initially thought that this conception represents the *nature* of human beings, but then he recognized that the interpretation of human nature is a controversial issue and that everyone would have their own answer after rational deliberation. Not all people understand themselves in

⁸ While some philosophers interpret that Rawls limits the application of his principles within liberal democratic societies, some argue that Rawls does not limit the application of his principles but simply take the public political culture of such societies as the place to draw the fundamental ideas for justice as fairness. I cannot compare these two interpretations in detail here. Moreover, which one is more correct is not a very important question since, even the later interpretation is more correct, my argument would not be affected. For an example of the earlier interpretation, see Buchanan (2000: 80).

⁹ According to Freeman, Rawls did not give up this comprehensive doctrine throughout his entire life. See Freeman (2007c: 327).

this Kantian way. Therefore, in his later writings, Rawls gives up the claim that people are by nature free and equal persons. Rather, he argues that members of a liberal democratic society share several political conceptions, such as ‘the idea of citizens as free and equal persons’ and ‘the idea of a well-ordered society as a society effectively regulated by a political conception of justice’.¹⁰ Rawls believes that, provided that a person grew up under the public political culture of a liberal democratic society, he would acquire these political conceptions. These political conceptions represent a common perspective shared by the members of liberal democratic society.

However, those who live in non-liberal-democratic society have a different public political culture, hence they might share different conceptions of the person as well. Because of this difference, political conceptions may not be generally acceptable in these kinds of society. In light of the relationship between conceptions of the person and public political culture, the later Rawls recognizes that his social contract has a limited scope and would only be generally accepted in societies in which a specific kind of public political culture exists.

The motivation of contractors

After limiting the scope of the contract within a liberal democratic society, Rawls argues that the political conceptions represent a way of practical reasoning publicly shared among the members in this kind of society. The purpose of Rawls’ hypothetical contract is to discuss what principles would be agreed insofar as all people reason in this particular way. The hypothetical contractors have two moral

¹⁰ Rawls (1993: 14).

capacities: the capacity for a conception of the good and the capacity for a sense of justice, which correspond to the two aspects of practical reason respectively.¹¹

For the sense of justice, Rawls does not make any change of definition but only emphasizes that this desire is compatible with different views of human nature. The sense of justice is ‘the capacity to understand, to apply, and to act from the public conception of justice which characterizes the fair terms of social cooperation’.¹² According to Rawls, this is the desire to embrace principles which would be selected in the Original Position, since the situation represents the agreement that would be made when all parties were fairly situated. In *A Theory of Justice*, the sense of justice seems to require people to consider one another free and equal persons. This assumption guarantees everyone’s equal moral status.¹³ However, this metaphysical assumption of human nature disappears in *Political Liberalism*; now the sense of justice requires people to take everyone as free and equal only in a political sense. It is as if they play the roles of free and equal persons. They can have different views of human nature, but in the political domain they have to *act* as free and equal citizens and respect one another as such. Being a free and equal citizen is only ‘acting a part in a play, say of Macbeth and Lady Macbeth’.¹⁴

For the conception of the good, Rawls also gives up the view that the desires to realize the two moral capacities are the ‘highest-order’ interests. In *A Theory of Justice*, because of the Kantian interpretation of human nature, the two moral capacities characterize the fundamental nature of human beings, so realizing them is to the highest intrinsic good of everyone. However, in *Political Liberalism*, Rawls merely defines the capacity for a conception of the good as ‘the capacity to form, to

¹¹ Rawls (1993: 52).

¹² Rawls (1993: 19).

¹³ Rawls (1999a: 441-443).

¹⁴ Rawls (1993: 27).

revise, and rationally to pursue a conception of one's rational advantage or good'.¹⁵ Free and equal citizens could form different conceptions of the good, given that these conceptions were formed in light of formal principles of rational choice.¹⁶ In these conceptions of the good, the desires to realize the two moral capacities are not necessarily the highest-order interests. In fact, in his later writings, Rawls changes the name of these desires to 'higher-order' interests or 'fundamental' interests, but not 'highest-order' interests.¹⁷ This is not only a change in name, but also represents a change in the status of these desires. The later Rawls acknowledges that some free and equal citizens might take the realization of two moral capacities as only instrumentally good.¹⁸ They might not, as the earlier Rawls expected, take this as the highest intrinsic good in their life. But still, these interests are fundamental since people could only take part in social cooperation after they realize these two capacities.

In short, except for some minor differences, the characteristics of free and equal citizens are highly similar to those of free and equal persons in the earlier Rawls' writings. The sense of justice no longer presupposes a Kantian interpretation of human nature, whereas the conception of the good no longer presupposes that people have the highest-order interests in realizing the two moral powers. However, one should not easily overlook these minor differences. The presuppositions that Rawls gives up belong to the Kantian comprehensive doctrine, which is controversial even in the liberal democratic societies.

¹⁵ Rawls (1993: 19).

¹⁶ Rawls (1993: 176n. 3).

¹⁷ For the former change, see Rawls (1993: 74, 76, 107, 178); for the later change, see Rawls (2001: 104-106).

¹⁸ Rawls (2001: 104) and Freeman (2003: 307).

7.2.2 Hybrid contractarianism and the condition of priority

Despite minor differences, free and equal citizens would also choose principles of justice as fairness, as would free and equal persons. But can these principles fulfil the conditions of priority and generality? I will discuss how they satisfy the condition of priority in this section.

Similar to the earlier Rawls, the later Rawls achieves the condition of priority by showing the congruence between rationality and reasonableness. However, Rawls does not intend to prove this congruence in all contexts.¹⁹ Rather, he aims only to show that rationality and reasonableness are congruent with each other in a particular context characterized by political conceptions. As Rawls argues,

Just as the principles of logic, inference, and judgment would not be used were there no persons who could think, infer, and judge, the principles of practical reason are expressed in the thought and judgment of reasonable and rational persons and applied by them in their social and political practice. Those principles do not apply themselves, but are used by us in forming our intentions and actions, and plans and decisions, in our relations with other persons. This being so, we may call the conception of society and person ‘conceptions of practical reason’: they characterize the agents who reason and they specify the context for the problems and questions to which principle of practical reason apply.²⁰

Hence, political conceptions represent the context to which the rules that govern practical reason apply. Rationality and reasonableness have different implications in different contexts. For example, the rational choice in a religious context is different from a rational choice in a family context. As a Catholic, the most rational choice is

¹⁹ Rawls acknowledges that this point is not discussed clearly in *A Theory of Justice*. See Rawls (2001: 186).

²⁰ Rawls (1993: 107).

to seek for salvation; as a father, the most rational choice is to protect your family members and give them a good life. Rawls limits his ambition and merely consider *what principles of justice would be rationally and reasonably justified in the context characterized by political conceptions*. This context is about the use of coercive power in the basic structure; people are characterized by certain features and relationships in this domain. For example, in this context, a person ‘is seen as a free and equal citizen, the political person of a modern democracy with the political rights and duties of citizenship, and standing in a political relation with other citizens’.²¹ All citizens understand that the basic structure possesses political power, which is ‘always coercive power backed by the state’s machinery for enforcing its laws’.²²

Rawls then discusses which principles would be rationally and reasonably agreed and become an overlapping consensus insofar as all people adopt this common perspective, and he uses the hypothetical contract as the heuristic device. By identifying themselves as free and equal citizens, actual people should understand that they have reasons to honour principles of justice as have an overriding authority in the political domain, no matter what attitudes people have to principles of justice in other domains.²³ Since his argument for priority is similar to the argument that he used in his earlier writings, I will only briefly discuss it here. The later Rawls continues to adopt a dual conception of practical reason and thus defines free and equal citizens as having two moral capacities. By studying the choices of these citizens, actual people can know why principles of justice are both rationally and reasonably justified.

²² Rawls (2001: 182).

²³ Rawls (2001: 182)

From the reasonable perspective, principles of justice represent principles that could be publicly justified to each citizen. The later Rawls continues to use the Original Position to explain the reasonable justifiability of principles of justice. In the Original Position, the influence of arbitrary contingencies were excluded and each of the parties was treated fairly, so principles chosen in this situation would be principles that specify fair terms of social cooperation. As Rawls argues, these principles form ‘a basis of mutual respect’ and reasonable citizens who desire a relationship of mutual respect, should be motivated to embrace them.²⁴ ‘When they believe that institutions or social practices are just, or fair (as specified, say, by principles they would themselves, when fairly represented, be prepared to propose or to acknowledge), citizens are ready and willing to do their part in those arrangements provided they have sufficient assurance that others will also do theirs’.²⁵

From the rational perspective, principles of justice are justified also because they can protect fundamental interests that are highly valued by free and equal citizens. Although Rawls no longer proposes that the interests in realizing the two moral powers are the highest-order interest, these interests are still fundamentally important with respect to participating in social cooperation. Hence, from the perspective of free and equal citizens, everyone should choose principles of justice because these principles can guarantee that these fundamental interests are adequately satisfied.

Hence, in the political domain, free and equal citizens have reason to endorse principles of justice as fairness from both rational and reasonable perspectives. As I discussed in the last chapter, the key to satisfying the condition of priority is to incorporate both rationality and reasonableness into the conception of practical reason. Since Rawls preserves the dual conception of practical reason in his latter

²⁴ Rawls (1993: 303).

²⁵ Rawls (2001: 196).

writings, his latter contractarian theory still possesses the virtue of fulfilling the condition of priority. Insofar as there are no other aspects of practical reason except rationality and reasonableness, free and equal citizens can be seen as fully modelling the practical reasoning of actual people in the political context. Given that actual people conceive of themselves as free and equal citizens in the political context and principles of justice are both rationally and reasonably justified to free and equal citizens, actual people have no reason to object to these principles.

7.2.3 Hybrid contractarianism and the condition of generality

Rawls believes that, by making the ‘political turn’, his contractarian theory can satisfy the condition of priority without sacrificing the condition of generality. After showing that rationality and reasonableness are congruent in the context characterized by political conceptions, Rawls further shows that these political conceptions are generally acceptable. Compared with Hobbesian and Kantian contractarianism, Rawls’ assumption is much less general since he aims only to provide a social contract for liberal democratic society. However, Rawls argues that actual people in the liberal democratic society are generally willing to deliberate in terms of political conceptions. Rawls gives two reasons to explain the general acceptance of these political conceptions.

First, political conceptions are implicit in the public political culture of the liberal democratic society, thus actual people who live under this culture are all familiar with them. Although people belong to different associations and families and have different comprehensive doctrines, they all acquire these conceptions from the public culture. These political conceptions act as a shared approach to engaging in practical reasoning in public discussion. No matter what comprehensive doctrines people have, they are all able to use these conceptions as the ‘basic units of thought’ in the

political domain.²⁶ As Rawls argues, the public political culture is ‘the shared fund of implicitly recognized basic ideas and principles’.²⁷

The public culture comprises the political institutions of a constitutional regime and the public traditions of their interpretation (including those of judiciary), as well as historical texts and documents that are common knowledge...In a democratic society there is a tradition of democratic thought, the content of which is at least familiar and intelligible to the educated common sense of citizens generally. Society’s main institutions, and their accepted forms of interpretation, are seen as a fund of implicitly shared ideas and principles.²⁸

Secondly, these political conceptions are freestanding in that they are compatible with a wide range of comprehensive doctrines. Freestandingness means that a conception does not presuppose any comprehensive doctrine. Comprehensive doctrines, such as utilitarianism and the Kantianism that the earlier Rawls adopted, are doctrines that include ‘concepts of what is of value in life and gives life its meaning’.²⁹ ‘It covers the major religious, philosophical, and moral aspects of human life in a more or less consistent and coherent manner’.³⁰ In contrast to these comprehensive doctrines, a political conception is one ‘that formulates its values independent of non-political values and of any specific relationship to them’.³¹ The political conceptions merely represent a particular approach of practical reasoning and this approach is not derived from any specific comprehensive doctrines. They are ‘presented as freestanding and expounded apart from, or without reference to, any

²⁶ Rawls (1993: 18n. 20).

²⁷ Rawls (1993: 8).

²⁸ Rawls (1993: 14).

²⁹ Rawls (1993: 13).

³⁰ Rawls (1993: 59).

³¹ Rawls (1989b: 483).

wide background'.³² Because of their independence, most of the comprehensive doctrines are compatible with these ideas. Rawls writes,

A distinguishing feature of a political conception is that it is presented as freestanding and expounded apart from, or without reference to, any such wider background. To use a current phrase, the political conception is a module, an essential constituent part, that fits into and can be supported by various reasonable comprehensive doctrines that endure in the society regulated by it.³³

Moreover, Rawls reminds us that most of the comprehensive doctrines are not *fully* comprehensive. They usually have 'a certain looseness'.³⁴ Because of this looseness, there are many ways for the freestanding political conceptions 'to cohere loosely with those (partially) comprehensive views'.³⁵ People would accommodate these political conceptions into their comprehensive doctrines in different ways. Taking the political conception of free and equal citizens as an example, religious comprehensive doctrines would accept this political conception of person because principles of justice derived from this political conception can 'lead to a principle of toleration and underwrite the fundamental liberties of a constitutional regime'.³⁶ This political conception of the person can also be derived from the ideal of autonomy in Kantian and Millian comprehensive liberal doctrine.³⁷ Although the approaches of justification are different, all citizens can find reasons to accept the political conception of the person from their own comprehensive doctrines.

³² Rawls (1993: 12)

³³ Rawls (1993: 12).

³⁴ Rawls (1993: 159) and Rawls (2001: 197-198).

³⁵ Rawls (1993: 160).

³⁶ Rawls (1993: 145).

³⁷ Rawls (1993: xlv-xlv, 200).

In conclusion, the contractarian theory based on political conceptions overcomes the problem of generality that troubled the earlier Rawls. According to the later Rawls, the earlier version of justice as fairness is a set of principles that could not hope to order society without oppressive state action.³⁸ So he emphasizes that his assumption in his later theory is ‘political’ but not ‘comprehensive’. The political conceptions are generally accepted for two reasons. First, they are latent ideas implicit in the public culture that everyone has a chance to acquire. Secondly, they are freestanding conceptions that would not conflict with most of the comprehensive doctrines. Thus, principles of justice justified by these conceptions can act as the focus of an overlapping consensus among free and equal citizens who have conflicting comprehensive doctrines. Rawls is confident these political conceptions can be ‘an open and public basis of justification for citizens as free and equal’ within the liberal democratic society.³⁹ At least in a liberal democratic society, justice as fairness represents a social contract which can satisfy both the condition of priority and the condition of generality.

7.3 The problem of generality revisited

The ‘political’ turn seems to help Rawls’ hybrid contractarianism justify political principles which are both general and overriding. Rawls no longer presupposes the Kantian comprehensive doctrine and argues that, as long as citizens live in a just society, the freestanding, non-comprehensive political conceptions can be easily built into their comprehensive doctrines. However, I would like to argue that Rawls is still too optimistic about the general acceptability of political conceptions. Rawls expects that all citizens will use political conceptions as the only approach to engaging in practical reasoning in the political domain, but I believe that, even in a well-ordered

³⁸ Rawls (1993: xvi, 37).

³⁹ Rawls (1993: 115).

society, citizens are still ambivalent about the role that political conceptions should play in practical reasoning. Only a limited number of citizens will embrace principles of justice wholeheartedly as Rawls describes.

7.3.1 The limit of scope: from the world to liberal democratic societies

Strictly speaking, the later Rawls' social contract fails to fulfil the condition of generality since its assumptions could, at best, be generally acceptable only to members of liberal democratic society. Rawls bites the bullet and acknowledges that his conception of practical reason cannot be universally accepted. In this sense, Rawls' contractarianism is less general than Hobbesian and Kantian contractarianism. Hobbesian and Kantian contractarians expect their conceptions of practical reason can be applied to all human beings, regardless of their culture and society. Given that one is a human being, one is able to assess rationality or reasonableness. However, Rawls believes that his political conception of free and equal citizen represents a specific way of applying rationality and reasonableness in a certain context.⁴⁰ This specific way of reasoning can be acquired only by living in a liberal culture. For example, if some people are not members of liberal democratic society, then they might simply have no interest in rationally realizing their two moral capacities.

Because of its limited applicability, Rawls' hybrid contractarianism becomes less appealing. It loses one of the virtues of contractarianism. In Chapter 2, I showed that one of the virtues of contractarianism is that it can justify the authority of political principles even in a highly diversified society. A modern society is usually characterized by a plurality of perspectives on political arrangements. It is unlike the

⁴⁰ Rawls (1993: 107-109).

kind of small, consolidated community which has a substantial political consensus. Different people have different views and these views inevitably conflict with one another. Thus contractarians rely on some uncontroversial conceptions of practical reason, such as the instrumental conception of rationality adopted by Hobbesian contractarianism, to derive political principles which are substantial but still publicly justified.

However, the later Rawls' contractarianism can be applied only in a society that already has an antecedent agreement on several political conceptions. The conception of justice as fairness is a contingent, overlapping consensus among people with varying views. This undermines the significance of hybrid contractarianism, since its justification is directed to a specific group of people but not, like Hobbesian and Kantian contractarianism, directed to an indefinitely large circle. It must presuppose an antecedent agreement and only people who are in this agreement are the audience for justification. This means that Rawls' contractarianism is helpless in a highly pluralistic society.

One might argue that even though Rawls' contractarianism can be applied only in a specific kind of society, it is already an impressive achievement in political philosophy.⁴¹ In fact, apart from Rawls, it is not uncommon for philosophers to think that the demands of rationality and reasonableness can only be compatible with each other under certain social conditions. For example, Thomas Nagel, who also cares about this conflict (in his terminology, this is the clash between morality and good life), argues that one of the possible approaches to avoid this conflict is to live under proper political institutions. 'An important, perhaps the most important task of political thought and action is to arrange the world so that everyone can live a good

⁴¹ Burton Dreben holds this point. See Dreben (2003: 327-329).

life without doing wrong, injuring others, benefiting unfairly from their misfortune, and so forth'.⁴² Similarly, Samuel Scheffler also argues that a good society shapes the psychological structure of citizens, so that citizens can avoid the conflict between rationality and reasonableness. Proper social institutions 'nurture the psychological bases of effective moral motivation' and 'reduce the degree and frequency of conflict between moral requirements and the interest of the agent'.⁴³ The conflict between rationality and reasonableness is so deep that it seems most philosophers would agree that it can only be avoided in a particular kind of society. Therefore, it is too demanding to think that Rawls' contractarianism should be generally acceptable in all societies. Given the difficulty in reconciling rationality and reasonableness, we should appreciate Rawls' accomplishment since he demonstrates that, at least in a liberal democratic society, justice as fairness represents a social agreement that is both rationally and reasonably justified.

7.3.2 The limit of scope: from real liberal democratic societies to ideal well-order society

We can take a more sympathetic view of Rawls' project and understand that Rawls' social contract is limited only within the liberal democratic societies. Yet, I believe that Rawls is still too optimistic about the general acceptability of political conceptions. The congruence between rationality and reasonableness depends on freestanding political conceptions, which represent a specific, political way of engaging in practical reasoning.⁴⁴ By deliberating in terms of these political conceptions, rationality and reasonableness can be reconciled and the condition of priority can be fulfilled. Rawls then expects the political conceptions to be generally

⁴² Nagel (1986: 206).

⁴³ Scheffler (1992: 139).

⁴⁴ Rawls (1999c: 87).

acceptable to citizens and to act as a common ground for public justification. Rawls is no doubt correct that these political conceptions are more easily acceptable after being detached from the Kantian comprehensive doctrine. However, these political conceptions are still too exclusionary to be the basis of overlapping consensus among comprehensive doctrines in the liberal democratic society.

For example, some comprehensive religious doctrines, such as the modern Roman Catholic doctrine, are clearly incompatible with these political conceptions. This Catholic Church is in fact the largest church in the United States.⁴⁵ The population that believes this comprehensive doctrine is large.⁴⁶ The Catholic doctrine and the Rawlsian political conceptions have irresolvable epistemological disagreements; Rawls presupposes that free and equal citizens all accept the burdens of judgment. They should accept the fact that all people would answer some fundamental questions, such as questions of religion or truth, differently. Since they use their power of reason in different ways because of their life experience, they should arrive at different conclusions. The ultimate answers about religion and truth are hard to think through even under the best circumstances. However, the Roman Catholic doctrine obviously has a different understanding. For ‘a religious doctrine—as a purportedly authoritative guide to moral requirements and/or salvation—characteristically presents itself as universally accessible to clear minds and open hearts’.⁴⁷ It believes that, given that people could reason correctly, they should find that the Catholic doctrine is true. ‘Heresy and infidelity are due to worldly temptation, demonic intervention, divine predestination, and so on’.⁴⁸ From this example, we can see that even though Rawls believes that the political conceptions are freestanding enough, they would still conflict with some comprehensive religious

⁴⁵ Wenar (1995: 81n. 16).

⁴⁶ Klosko (2009: 30). For an elaboration of this point, see Klosko (2000: 150-182).

⁴⁷ Wenar (1995: 66).

⁴⁸ Wenar (1995: 66).

doctrines that are prevalent in the liberal democratic society. As George Klosko says, Rawls' description of liberal democratic society beautifies the reality too much, since the existing liberal democratic societies are in fact unstable and full of conflicts.⁴⁹ This is also why Joseph Raz comments that 'Rawls' route seems barren in pluralistic societies....The degree of existing diversity is just too great'.⁵⁰

One might defend Rawls by arguing that the assumption of the burdens of judgment is merely a dispensable element in the political conceptions, and one that is incompatible with these comprehensive religious doctrines. The problem can be avoided by excluding the burdens of judgment from the political conceptions.⁵¹ But another more plausible response would be that Rawls' contractarianism should not be misunderstood as arguing for political principles which 'can be justifiable to all (or most) actual citizens in liberal democracies, even those people who do not accept certain basic liberal norms'.⁵² Rather, the aim of Rawls is to understand how political principles 'can be publicly justified to the constituency of an ideal liberal democratic society'.⁵³ Since Rawls' social contract is an ideal theory, it is unfair to criticize it by

⁴⁹ Klosko (1994: 1896).

⁵⁰ Raz (1990: 425). Nevertheless, in this article, Raz seems to overlook the fact that Rawls is actually proving that justice is fairness could receive the hypothetical consent from rational and reasonable citizens who already share political conceptions, but not proving that it could receive actual consent of citizens in the wider world.

⁵¹ This is the point suggested by Leif Wenar. See Wenar (1995: 76-78). In fact, from our previous analysis of Scanlon, we can also see that the presupposition of the burdens of judgments is unnecessary in the conception of reasonableness. Scanlon's conception of reasonableness, which is the conception of practical reason that can fulfil the condition of generality, only presupposes that people have a willingness to behave in a way which could be justified to one another. Adding the epistemological presupposition of the burdens of judgment to the conception of reasonableness would only undermine the general acceptability of this conception.

⁵² Quong (2011: 5).

⁵³ Quong (2011: 6). Jonathan Quong also offers a comprehensive discussion to explain why we should not understand the later Rawls' contractarianism (Political Liberalism) as a theory that aims at providing justification 'to the diverse constituency of persons that currently inhabit modern liberal societies.' See Quong (2011: 137-160).

the unreasonable doctrines existing in the real world.⁵⁴ Maybe it is true that, as some theorists suggest, because of the realistic political limits, we can at most expect a *modus vivendi*.⁵⁵ However, the task of Rawls to show that if actual people lived in the perfect, ideal, well-ordered society presented by Rawls, then they would be possible to share political conceptions and attain a social contract that could be rationally and reasonably justified. Those critics who argue that political liberalism is inapplicable to the existing diversified world misunderstand the real ambition of Rawls.

7.3.3 Weakly just citizens and strongly just citizens

Although Rawls limits the scope of his social contract to the well-ordered society, I will argue in this section that such an ideal social circumstance is unsustainable and political liberalism would eventually fail to be publicly justified. First we have to understand the features of this ideal society. According to Rawls, a well-order society is, ‘first...a society in which everyone accepts, and knows that everyone else accepts, the very same principles of justice; and second, its basic structure...is publicly known, or with good reasons believed, to satisfy these principles. And third, its citizens have a normally effective sense of justice and so they generally comply with society’s basic institutions, which they regard as just.’⁵⁶ Because of these special characteristics, the pluralism in this ideal society is ‘reasonable pluralism’, but not the normal pluralism that is commonly seen in the existing world. Pluralism is ‘simply a variety of doctrines and views, as one might expect from people’s

⁵⁴ Rawls (1993: 55). However, this interpretation goes against some theorists’ observation. For example, Adam Swift and Stephen Mulhall believe that the burden of judgments is a sociological observation but not philosophical claim. See Mulhall and Swift (1996: 224).

⁵⁵ Dauenhauer (2000: 213-219).

⁵⁶ Rawls (1993: 35).

various interests and their tendency to focus on narrow points of view'.⁵⁷ Reasonable pluralism is different from this, in that it is formed by 'a diversity of reasonable comprehensive doctrines'.⁵⁸ Although citizens still hold doctrines which are different from one another, their doctrines are compatible with political conceptions and citizens have a shared commitment to give political conceptions priority in their practical reasoning.

The well-ordered society is no doubt an idealized conception which does not refer to any society in the real world. Given such an ideal society, what kind of political principles would be publicly justified? What social contract could be formed among citizens in the well-order society? These are the questions that Rawls' contractarianism aims at answering. Rawls argues that, despite the fact that citizens have different comprehensive doctrines in this society, principles of justice in political liberalism would still be rationally and reasonably justified to each citizens. These principles represent a public social contract that could be legitimately imposed in this society. Hence, political liberalism does not aspire to justify principles of justice to actual people in the real would. Rather, the group to which it offers justification is the hypothetical citizens in the well-ordered society, who already share the political conceptions and respect them as the common ground of public justification.⁵⁹

⁵⁷ Rawls (1993: 36).

⁵⁸ Rawls (1993: 36).

⁵⁹ This is what Jonathan Quong calls 'internal conception of political liberalism'. Quong argues that political liberalism should not be understood as an 'external conception'. It does not aim at showing that a liberal conception of justice can be justified to non-liberal citizens who exist in the actual liberal democratic societies and do not accept certain basic liberal norms. Rather, the aim of political liberalism is to 'understand what kind of arguments, if any, citizens *already committed* to certain basic liberal norms can legitimately offer to one another'. (Quong 2011: 5) It is an internal conception because it only shows that the political conception of justice is internally consistent with reasonable

However, it does not mean that political liberalism is merely a hypothetical contract irrelevant to the real world. Although political liberalism refers to a hypothetical contract in an ideal society, it is realistically possible (this is why Rawls calls this ideal society ‘realistic utopia’). Rawls explains that political liberalism could be realized given that a specific kind of citizens was created. Insofar as actual people were brought up in the well-ordered society, they could learn to be free and equal citizens and develop reasonable comprehensive doctrines.⁶⁰ This possibility is explained by the reasonable moral psychology. Rawls distinguishes the whole process of transformation into three stages.⁶¹ The first stage is called ‘the morality of authority’. Children should be brought up under the care of their parents. Lovingly cared for and nurtured by their parents, they are able to acquire new emotions. They gradually develop a loving attachment to their parents and this attachment is independent of their self-interest.⁶² The second stage is ‘the morality of association’. When the children grow up, they inevitably enter into different associations. Provided that these associations are just and other fellow associates ‘live up to their duties and obligations, [children] would develop friendly feelings toward them, together with feelings of trust and confidence’.⁶³ These feelings for their fellows are also new feelings independent of people’s self-interests. People become free and equal citizens in the third stage when they come to embrace ‘the morality of principle’. They recognize the existence of principles of justice and ‘understand the value they secure and the way in which they are to everyone’s advantage’.⁶⁴ Principles of justice make the families and associations that they value possible.

pluralism in the well-ordered society. For a detailed discussion of the internal and external conception of political liberalism, see Quong (2011: 137-159).

⁶⁰ Rawls (2001: 195-196).

⁶¹ This three-stage development was first mentioned in *A Theory of Justice*. Although it was rarely mentioned afterward, Rawls continues to take it as a part of his theory even in his later writings. See Rawls (2001: 196).

⁶² Rawls (1999a: 406). A similar observation can also be found in Rousseau (1979: 174).

⁶³ Rawls (1999a: 411).

⁶⁴ Rawls (1999a: 414).

Because of this moral development, citizens can be developed into a specific kind of moral agent and the well-ordered society is sustainable.

Even though citizens can acquire political conceptions through the three-stage process, I believe that Rawls is too optimistic that all the citizens in the well-ordered society will come to hold reasonable comprehensive doctrines. We should distinguish two kinds of citizens here: *weakly just citizens* and *strongly just citizens*. For weakly just citizens, their comprehensive doctrines include political conceptions. They know how to conceive of themselves and of society in this particular ‘political’ way. Political conceptions affect how they look at themselves and their society. But apart from having political conceptions, their political vision is also affected by other non-political conceptions. They think that they are *not only* free and equal citizens and society is *not only* shaped by fair terms of social cooperation. They still have other political ambitions and they will reject political principles when these principles overlook some non-political conceptions.⁶⁵ Strongly just citizens will not only accept political conceptions, but also respect them as the sole ground in the justification of political principles. The perspective of free and equal citizens is the *only* perspective that they would use in the public domain. When they think about what political principles should be justified, they put other considerations in their comprehensive doctrines aside. In the public domain they consider justification merely in terms of political conceptions.

Clearly the stability of the overlapping consensus that Rawls expected can only be maintained by strongly just citizens, since this kind of citizen generally accepts all

⁶⁵ Some might criticize that the name of weakly just citizen is misleading because if they are only willing to act in accordance with the principles of justice when this does not conflict with their own aims, then they are in fact not just. I will discuss this question in detail in Section 7.3.4. Moreover, I am concerned with the question of whether this kind of citizens is possible to emerge in a well-ordered society, more than the question of whether these citizens are just.

the assumptions of justice as fairness and is always willing to endorse justice as fairness, which is justified by political conceptions. Only the doctrines held by these citizens qualify as ‘reasonable comprehensive doctrines’.⁶⁶ Rawls also believes that, given that citizens live in a well-ordered society, they can become strongly just citizens.⁶⁷ However, I believe that the ideal circumstances of a well-ordered society can cultivate only weakly just citizens. Hence the overlapping consensus is unsustainable.

Suppose a person was raised in a well-ordered society. In this person’s experience of living with his family and various associations, he is loved and fairly treated by his family members and fellow associates. His development is similar to what Rawls describes. He starts to respect justice as fairness and acquire the political perspective of a free and equal citizen. But one day he passes a bookstore and finds Marx’s *Kapital*. He immediately becomes a follower of Marx, finding his theory highly appealing; a proletarian revolution should be brought about, the bourgeoisie should be overthrown and private property should be abolished.⁶⁸ He takes a socialist world to be the most ideal world for it can lead all people to develop their nature freely in cooperative production without exploitation. Hence, quite apart from the perspective of the free and equal citizen, he acquires a new political vision. It does not mean that he finds that justice as fairness is unappealing and political conceptions should be rejected. He is different from the Catholics that we discussed before, who do not accept the political conceptions. Because of his past, he still values political conceptions, such as the idea of free and equal citizens and the idea of fair terms of social cooperation (supposing that he has some way to accommodate the conflicts between these schools). However, he now believes that these ideas are only part of

⁶⁶ Rawls (1997: 591-592, 608-609).

⁶⁷ Rawls (2001: 196-197).

⁶⁸ As Freeman argues, justice as fairness is compatible with a property-owning democratic institution. See Freeman (2007c: 28). See also Rawls (2001: 158-162).

the things that should be considered in justifying political principles. As a socialist, he believes that justice as fairness is inadequate since citizens living in this society still cannot avoid the alienation of labour. Fully justified political principles should also be justified from a socialist perspective. Merely taking political conceptions into account is inadequate.

This socialist is a kind of weakly just citizen. In fact, a socialist case like this is not uncommon in a well-ordered society. Under the protection of freedom of expression and freedom of association, citizens can exercise their reason in countless ways. Rawls surely acknowledges the diversity caused by the free exercise of human reason.⁶⁹ But he also believes that, provided that citizens live in a well-ordered society, they will become strongly just citizens who would not be affected by these non-political values in the public domain. However, I believe that the result of the free exercise of human reason is more complex than Rawls suggests and, because of the free exercise of human reason, we can at most expect that people develop into weakly just citizens in the well-ordered society. Rawls is correct that, from the love and fair treatment of family and fellow associates, children would learn political conceptions and would become free and equal citizens. But they would also get in touch with different ideas and would develop different kinds of affective ties with other people. These ideas and affective ties would influence their comprehensive doctrines. All these comprehensive doctrines have their own ways of engaging in practical reasoning in the political domain, and these ways may not necessarily be compatible with justice as fairness. As Kukathas argues, insofar as a liberal government guarantees that people can exercise their reason freely, non-liberal citizens inevitably emerge.

⁶⁹ Rawls (1993: 37).

Liberal political systems...accept within the polity people who would like to see that political system abandoned: anarchists, religious fundamentalists, communists, ethnic nationalists, and fascists among others. In liberal societies such people are free not only to hold to such views but also to proselytize: to argue against democracy and free speech, to discourage people from voting, to run for office on an anti-liberal ticket with the endorsement of an anti-liberal party, and even to write and publish books excoriating liberalism.⁷⁰

Hence, the three stage moral development only explains how these political conceptions are built into the doctrines, but does not explain why these political conceptions can have a special status in the justification of political principles. The citizens in a well-ordered society should have more than one approach for engaging in practical reasoning in the political domain. If political conceptions only represent one of these approaches, why must citizens take this approach as the only approach in the justification of political principles?

In fact, this problem is more obvious when other perspectives of understanding politics are closely related to some fundamental commitment in citizens' comprehensive doctrine. Rawls thinks that a comprehensive doctrine determines 'what is of value in human life and ideals of personal character'.⁷¹ These values and ideals give meaning to the lives of citizens. If these values and ideals imply certain approaches of engaging in practical reasoning in the political domain, giving up these approaches means that they have to betray their fundamental commitments. So why should citizens give up these approaches and deliberate in terms of political conceptions only?⁷² Taking the socialist as an example again: given that he firmly believes that some of his fellow members of society are being exploited and the

⁷⁰ Kukathas (2001: 328).

⁷¹ Rawls (1993: 13).

⁷² In fact, Rawls also acknowledges that in a religious comprehensive doctrine, transcendental values such as salvation and eternal life are superior to the political conceptions. But if this doctrine is reasonable, then the political conceptions should still be able to override these transcendental religious values. However, Rawls does not explain *why* it is possible. See Rawls (1997: 609).

current state is an instrument of class domination, and his ultimate goal in life is to liberate his fellows from the government controlled by the capitalists, he has no reason to leave all these thoughts aside when he thinks about what political principles are justified.

This shows the limit of Rawls' reasonable moral psychology. Although children can develop into citizens who believe political conceptions, they may not accept that these conceptions are the sole considerations during the justification of political principles even they were brought up in the well-ordered society, since they may acquire other perspectives of understanding politics and treat political conceptions in different ways. There is no reason to believe that parental love and the fair treatment of associations has a larger effect on the lives of citizens than other influences. Citizens are possibly more affected by other influences and may develop comprehensive doctrines that do not take political conceptions as the only way of justifying political principles. Although Rawls recognizes the diversity caused by the free exercise of human reason, he still underestimates it. Rawls' description of a reasonable moral psychology is not incorrect, though it is, in the end, *only one of the possible* ways in which citizens might develop in the well-ordered society.

7.3.4 The tension between freestandingness and overridingness

Rawls might respond that the idea of a weakly just citizen is misleading because when a citizen accepts a political conception without taking it as the sole ground of public justification at the same time, then he is unjust. The respect for political conceptions is related to the reasonable capacity of a free and equal citizen.⁷³ One of the political conceptions that citizens learn in the course of their moral development

⁷³ Rawls (1993: 226).

is the political conception of free and equal citizens. After the three-stage process of moral development, citizens should be able to conceive of themselves as free and equal citizens who are both rational and reasonable. If they are reasonable, they should understand the burdens of judgment and that political principles should be publicly justified to each person. They would not propose principles which allow using political power to apply certain religiously true or philosophically true policies.⁷⁴ On the other hand, political conceptions are the only shared common ground among citizens in the well-ordered society. So if citizens want to justify political principles in a publicly acceptable way, then they can only use political conceptions and would need to put other ideas from their comprehensive doctrines aside. That is why political conceptions and the duty to respect these conceptions are inseparable.

However, justifying this duty by invoking political conceptions simply begs the question. The problem is that other ideas and values in citizens' comprehensive doctrines can also generate similar duties. These ideas and values can also justify a duty that citizens should consider other non-political conceptions during the justification of political principles and should not give public justifiability so much weight. For example, in the socialist's mind, he has a lofty obligation to fight for a socialist utopia, and this commitment generates a duty to spread socialist ideas even in the public domain. So why should citizens give the duty to respect political conceptions a special status? Given that political conceptions are less crucial than some non-political conceptions in citizens' comprehensive doctrines, citizens have no reason to give the duty generated by political conceptions—but not the duty generated by these non-political conceptions—a special status. In fact, if the duty to respect political conceptions comes from political conceptions themselves, it is hard to explain why political conceptions, which are merely some of the many ideas in

⁷⁴ Rawls (1997: 609).

citizens' comprehensive doctrines, can generate a duty which can outweigh other non-political conceptions that conflict with it.

From this we can identify a tension in the definition of political conceptions. On the one hand, Rawls defines political conceptions as freestanding, such that they are not connected with any comprehensive doctrine. They have nothing to say about some fundamental philosophical questions in human life, such as what is truth and what the meaning of life might be. Although they are moral ideas, they do not have any strong and comprehensive ethical implications.⁷⁵ On the other hand, Rawls says that political conceptions are able to establish a duty which is great enough to outweigh all the non-political conceptions. No matter how important these non-political conceptions are in the lives of citizens, citizens should still be strongly just and not be affected by them (or only affected by them to a limited extent).⁷⁶

However, it is unclear how political conceptions, which are merely freestanding, intuitive ideas that do not have any strong philosophical and ethical implication, can create a duty which can outweigh citizens' fundamental philosophical and ethical beliefs. Human beings usually have their own answers to fundamental philosophical and ethical questions. These philosophical or ethical beliefs, though non-political, provide accounts of meaning in life and affect how they understand the world. Since these beliefs are fundamentally important in people's life, people's actions which are caused by these beliefs would not be easily affected by other intuitive or fragmented beliefs. If people are willing to ignore these philosophical and ethical beliefs because of political conceptions, then these political conceptions inevitably must have something to say about the fundamental philosophical and ethical questions. But if

⁷⁵ Rawls (1993: 11).

⁷⁶ Rawls loosens his duty in his later writings. Even if citizens can appeal to non-political conceptions in the public domain, their way must be compatible with political conception. See Rawls (1997: 591-592).

the political conceptions are really connected with certain liberal comprehensive doctrines, then they will no longer be as freestanding as Rawls presents. Political conceptions are no longer a relatively simple framework that can facilitate public discussion among citizens of diverse backgrounds and faiths. Hence, the assumption of strongly just citizens requires Rawls to assign to political conceptions a comprehensive role that is incompatible with the freestanding nature of political conceptions. Justifying the duty to respect political conceptions by political conceptions themselves merely undermines the general acceptability of political conceptions.

In light of how rare strongly just citizens are, I believe that the scope of Rawls' hybrid contractarianism is much more limited than Rawls imagines. Compared with the earlier Rawls' contractarianism, the later Rawls' contractarianism is surely compatible with more comprehensive doctrines. Free and equal citizens could hold reasonable comprehensive doctrines which include a wide range of possibilities, such as liberal Catholicism, liberal utilitarianism, Kantianism, Millianism, value pluralism, and so on. However, as Thomas Hill indicates, the change in Rawls is actually not so radical, for most of the things in *A Theory of Justice* are preserved.⁷⁷ If many comprehensive doctrines are incompatible with the Kantian comprehensive doctrines of *A Theory of Justice*, then most of them will remain incompatible with the political conceptions. Many comprehensive doctrines can embrace political conceptions but do not treat political conceptions as the sole basis for justifying political principles. These comprehensive doctrines can surely emerge in a well-ordered society. As we saw before, the circumstances of a well-ordered society can at most guarantee that people become weakly just citizens, but cannot guarantee that they become strongly just citizens. The socialist doctrine is one of many possibilities that would appear in a well-ordered society. In the end, Rawls' political conception

⁷⁷ Hill (1994: 349). See also Wolf (2000: 103).

of justice would only be accepted among strongly just citizens, who are only *some* of the citizens in the well-ordered society.⁷⁸

It is worth noting that I am not criticizing the problem of partial compliance of Rawls' contractarianism. I am not saying that in the real world many actual people would disagree with Rawls' assumption or that modern society is so diversified that an overlapping consensus is unrealistic.⁷⁹ From the example of Roman Catholic Church, we can see that these kinds of people are inevitable. Yet this question is neither Rawls' concern nor mine. Since Rawls is proposing an ideal theory, he needs only to prove that, in the ideal social circumstances that he describes, it is realistically possible that citizens endorse justice as fairness.⁸⁰ Because of many problems in the real world, maybe an overlapping consensus on justice as fairness will never come to exist, but this political ideal is still realistically *possible*.⁸¹ In a well-ordered society governed by justice as fairness, this political ideal is stable and sustainable. There are many actual people who disagree now only because the current social circumstance affects their growing experience. Rawls made a political

⁷⁸ However, it does not mean that the political conception of justice as fairness is not consistent with reasonable pluralism. Suppose that there is a society in which only has strongly just citizens, the political conception of justice would be publicly justified. Nevertheless, this society is unsustainable because, insofar as citizens could exercise their reason freely, weakly just citizens would emerge and reasonable pluralism would disappear. The scope of social contract would then become limited within the strongly just citizens in this society.

⁷⁹ This is one of the most common misunderstandings of Rawls. Many philosophers think that Rawls tends to provide a *solution* of real politics; hence he underestimates the existing conflicts too much. But in fact Rawls only suggests a *possibility* that justice as fairness is realistically possible to be realized in the real world. For the example of this misunderstanding, see Raz (1990), Barry (1995b), Klosko (2000), Wenar (1995), Bohman (1995), Friedman (2000), Mills (2000), Wolf (2000) and Dauenhauer (2000).

⁸⁰ This is the central concern of Rawls. See Dreben (2003: 322)

⁸¹ Rawls (2001: 197) and Hill (1994: 336).

turn in his later writings because, even in an ideal well-ordered society, the Kantian comprehensive doctrine is still not generally accepted among citizens.⁸²

However, the critique I want to make is that, despite justice as fairness being less comprehensive, its assumptions are still too substantial in that it would not be widely adopted in a well-ordered society. Because of the free exercise of human reason and the free circulation of thought, it is not generally acceptable among people even in the most ideal circumstances. Weakly just citizens would emerge (and not in a minority) and would refuse to be one of the members of Rawls' social contract. Hence the problem of realistic possibility still remains. Rawls believes that his earlier expectation that the Kantian comprehensive doctrine would be widely adopted in a well-ordered society is unrealistic, but it seems that it is not less unrealistic to expect that a particular way of treating political conceptions would be widely adopted without being questioned by citizens.⁸³ If Rawls takes the controversy over the Kantian interpretation of human nature as a critical problem in his earlier writings, then he would be unable to deny that the controversy over the importance of political conceptions is a critical problem as well, since the two problems have the same nature.

One might defend for Rawls by arguing that this critique misunderstands Rawls' ambition. He merely aims at showing that it is realistically possible to have a well-ordered society in which his principles of justice could be publicly justified among

⁸² Rawls (1993: xviii-xix) and Freeman (2007c: 411-412).

⁸³ In fact, Rawls does not hold a radical standpoint that all unreasonable comprehensive doctrines can be eliminated in the well-ordered society. He believes that unreasonable comprehensive doctrines are inevitable, yet in the well-ordered society these doctrines 'do not undermine the unity and justice of society'. (Rawls 1993: xix, 60-61) However, if what I showed above is true, then the number of unreasonable comprehensive doctrines should be many more than Rawls expects. There are too many comprehensive doctrines which can contain political conceptions but merely take them as one of the many ways of understanding politics.

citizens. Even if the emergence of weakly just citizens is inevitable, Rawls's theory is still coherent. For when weakly just citizens occur, the society is no longer well-ordered, and it is not the kind of society with which Rawls is concerned.⁸⁴ However, I am not sure whether Rawls would satisfy with this level of achievement. As Rawls himself argues, a conception of justice is "seriously defective" if it fails to engender in human beings the requisite desire to act upon it.⁸⁵ He does not only want to show that a well-ordered society could possibly exist, but also wants to show that such a society is self-sustainable. The society can always produce the specific kind of citizens which is necessary for maintaining its existence. Citizens would not be tempted to be unjust and would support the just institutions continuously. Yet, as we saw before, Rawls overestimates the stability of a well-ordered society. Even in the ideal social circumstance of a well-ordered society, people would grow up in various ways and only some of them would become strongly just citizens. The conception of justice as fairness would be challenged by weakly just citizens and a well-ordered society would inevitably become unstable. Although Rawls's social contract becomes more freestanding and widely acceptable, it still falls prey to the problem of instability.

Interestingly, the philosophical problem that Rawls fails to solve also bothered Rousseau, who is another hybrid contractarian. In *Emile*, Rousseau argues that whether a civil city is stable depends on whether a particular kind of citizen can be created. Citizens can be educated properly insofar as they live in a society ruled by a Great Legislator. By acquiring correct knowledge, citizens will share the same communal selves and 'believe [themselves] no longer one but a part of the unity and no longer feel except within the whole.'⁸⁶ They will respect laws that can preserve their freedom and natural goodness. Thus a stable civil society can be created. In

⁸⁴ I am indebted to Jonathan Quong for this possible defence.

⁸⁵ Rawls (1999a: 398).

⁸⁶ Rousseau (1979: 40).

Rawls' terminology, these citizens share a conception of citizenship and always respect this public basis of justification.

However, as Judith Skhlar argues, Rousseau's project of creating citizens is problematic, because the emergence of individual selves is inevitable and these individual selves will conflict with the shared communal selves.⁸⁷ When children grow up, they naturally develop inner lives of their own and this is a necessary consequence of the free exercise of reason. These individual selves play a far more crucial role in people's lives than the communal selves do. Independence and self-esteem can only flourish if one's individual self is preserved. Insofar as the emergence of individual selves is natural, it is very difficult to create citizens who only have communal selves. This is why Rousseau is so deeply aware of the individuality of each person and argues that the Great Legislator should adopt a strict surveillance regime.⁸⁸ But Rawls cannot propose the same measures since the first principle of justice as fairness guarantees that everyone has equal basic liberties, which include freedom of speech. Given that the conditions of generality and priority can be satisfied insofar as a particular kind of citizens is created, and the creation of citizens involves totalitarian measures that Rawls would not accept, Rawls is unavoidably caught in the contractarianism's dilemma. The coercive measures proposed by Rousseau indicate that creating citizens is in fact not as easy as Rawls thinks.

In conclusion, even after the political turn, Rawls still fails to achieve the condition of priority without sacrificing the condition of generality. Since he aims at achieving the condition of priority he must presuppose that citizens are all strongly just citizens in order to explain why they would always deliberate in terms of political

⁸⁷ Skhlar (1969: 159-160).

⁸⁸ Skhlar (1969: 156-157).

conceptions in the political domain. Yet, as with the Kantian interpretation of human nature, this presupposition undermines the general acceptability of Rawls' contractarianism. It is not only that, in the real world, there are comprehensive doctrines like the Roman Catholic doctrine that are incompatible with political conceptions. Even in a well-ordered society in which every citizen was 'properly' brought up, they would only become weakly just citizens. Hybrid contractarianism is not prepared to deal with those citizens who are not reasonable enough: 'Political Liberalism does not engage those who think in this [unreasonable] way'.⁸⁹ However, the fact is that it is possible for many of these citizens who refuse to be members of Rawls' social contract to appear even in a well-ordered society.

7.4 Creating ideal citizens by education

In the last section, I showed that the general acceptability of hybrid contractarianism is still unsatisfactory despite Rawls' political turn. Rawls has two possible responses: one is about social circumstance and one is about philosophical reason. First, Rawls could emphasize the educational effect of the social circumstances of the well-ordered society. Secondly, Rawls could give philosophical reasons to explain why citizens in the well-ordered society should develop reasonable comprehensive doctrines. In Section 7.4 and 7.5 I will examine these two responses respectively.

7.4.1 Public political culture and background culture

One of Rawls' fundamental beliefs is that, when under proper social circumstances, children can be educated to become strongly just citizens in the well-ordered

⁸⁹ Rawls (1997: 574).

society.⁹⁰ Rawls emphasizes the educational influence of public political culture to citizens. Given that citizens live under a liberal democratic culture, they should be affected by this culture and should learn to be strongly just.⁹¹ As Rawls notes,

If citizens of a well-ordered society are to recognize one another as free and equal, basic institutions must educate them to this conception of themselves, as well as publicly exhibit and encourage this ideal of political justice. This task of education belongs to what we may call the wide role of a political conception. In this role such a conception is part of the public political culture: its first principles are embodied in the institutions of the basic structure and appealed to in their interpretation. Acquaintance with and participation in that public culture is one way citizens learn to conceive of themselves as free and equal, a conception which, if left to their own reflections, they would most likely never form, much less accept and desire to realize.⁹²

However, the educational effect of the public political culture is too vague to the extent that it is unclear how influential it can be in the growth of citizens. As we saw previously, due to the free exercise of human reason, citizens can develop various kinds of comprehensive doctrines. They are also affected by different cultures during their growth, and public political culture is only one of the sources of influence.

For example, as Rawls himself indicates, besides public political culture, a well-ordered society also has a 'background culture', which consists of 'the culture of churches and associations of all kinds, and institutions of learning at all levels, especially universities and professional schools, scientific and other societies'.⁹³ In the background culture, citizens can spread their comprehensive doctrines and persuade others by non-political conceptions. They do not have the responsibilities of

⁹⁰ Rawls (1997: 580)

⁹¹ In Rawls' earlier writings, he also discusses the relationship between the society and citizens' conception of person. See Scheffler (1978).

⁹² Rawls (2001: 56).

⁹³ Rawls (1997: 576n. 13).

limiting themselves to use political conceptions and to act as free and equal citizens. They can even criticize political conceptions and this distinction between public and private domain in the sphere of background culture.⁹⁴ According to Rawls, the public political culture and the background culture are compatible with each other. Suppose Rawls' description is true in a well-ordered society, yet how can Rawls be sure that, in the development of children, the influence of public political culture must be greater than that of their background culture? From the public political culture, children learn to be free and equal citizens and understand society as embodying fair terms of social cooperation. However, they can also learn other conceptions of the person and other perspectives on society and politics from the background culture.⁹⁵ The assumption of the educational effect of the public political culture seems to overlook the fact that the background culture can have the same level of influence on citizens as well.

7.4.2 Examining the influence of families and associations

Apart from the public political culture, families and associations also contribute to the moral development of children. As Rawls indicates, although principles of justice are not applied *within* families and associations, they are applied *to* these groups.⁹⁶ Families and associations have an obligation to ensure that children are raised to become ideal citizens; for example, they must love and fairly treat their children. The education of children should include teaching them about political conceptions so that they know the history of these conceptions and the role of these conceptions in the public culture. '[Children's] education should prepare them to be fully

⁹⁴ Rawls (1997: 576).

⁹⁵ This point is also supported by Bruce Brower, who argues that citizens in the well-ordered society must be given chances and liberties to explore lifestyles that are alternative to justice as fairness, and to consider and discuss argument from all points of view (including non-liberal views). See Brower (1994: 25).

⁹⁶ Rawls (1997: 588-589).

cooperating members of society and enable them to be self-supporting; it should also encourage the political virtues so that they want to honour the fair terms of social cooperation in their relations with the rest of society'.⁹⁷

Rawls is quite vague on the obligations that families and associations must assume. These obligations seem to be quite loose in that families and associations still enjoy a lot of freedom in teaching their young members. Some things are clearly prohibited, such as the abuse and neglect of children and the failure to expose them to citizenship education in order to ensure that children are loved and fairly treated.⁹⁸ So children can develop as Rawls describes. However, provided that families and associations fulfil these obligations, it seems that they are free to teach non-political conceptions to children. For example, they have the freedom to tell them bedtime stories, take them to cultural activities during holidays and so on. Rawls intends to avoid imposing harsh restrictions on the educational activities of families and associations. 'At some point society has to rely on the natural affection and goodwill of the mature family members'.⁹⁹ Yet, if the obligations are really so loose, it is hard to guarantee that children will become the strongly just citizens that Rawls expects. Because of the freedom that just institutions allow families and associations, they can transmit non-political values, including perspectives on politics different from the political conceptions. These teachings may have an enormous influence on the development of children, even more than the influence of public citizenship education. Therefore, even though children are loved and fairly treated and they have chances to be exposed to citizenship education, it does not mean that they will necessarily become strongly just citizens. These conditions are only necessary conditions, but not sufficient conditions.

⁹⁷ Rawls (2001: 156).

⁹⁸ Rawls (1997: 598) and Macedo (2000: 169-170).

⁹⁹ Rawls (1997: 598).

One might reply that the scenario presented above would only occur in a non-well-ordered society. If the society is really well-arranged, then all families and associations should hold reasonable comprehensive doctrines which support principles of justice for right reason, that is, ‘all those who affirm the political conception start from within their own comprehensive view and draw on the religious, philosophical, and moral grounds it provides’.¹⁰⁰ Given that they all endorse political conceptions, families and associations should have no reason to teach children something incompatible with political conceptions.

In such ideal circumstances, children can avoid corruption and will not become anything but strongly just citizens. However, it seems to paint too rosy a picture of a well-ordered society. This picture is problematic since it assumes that all citizens in a well-ordered society are strongly just. In my discussion above, I showed that a well-ordered society can only guarantee that all citizens are weakly just. Since these weakly just citizens have different comprehensive doctrines, they will have different approaches of teaching children. The diversity can be explained by the relationship between conscience and the education of children. The transmission of views to the next generation is in fact an important expression of conscience. The conscience of citizens is not only in their minds; they also want to exemplify it to their offspring and the young members of their associations.¹⁰¹ Citizens usually want to ensure that the conscientious views that they value most will not die out and can continue to be held by someone in the future.¹⁰² Thus most parents are highly concerned with how their children develop and whether they are influenced badly. Adult members of associations also want to transmit their comprehensive doctrines to their young members.

¹⁰⁰ Rawls (1993: 147).

¹⁰¹ I am indebted to Chandran Kukathas for discussing this point.

¹⁰² Fishkin (1992: 154).

In light of the importance of the education of children, one should understand that weakly just citizens will not easily concede the right to educate children. They are not satisfied with transmitting only political values to their children and expect to further transmit other non-political values as well, in order to let their offspring know the political visions that they think are valuable. This disagreement on the teaching approach is an unavoidable outcome of the free exercise of reason. If citizens have conflicting comprehensive doctrines because of their liberty of conscience (as we saw in 7.3.3), then they should have conflicting understandings of approaches to teaching as well. As Kukathas argues, everyone agrees that children should be educated properly, but what is disputed is what a proper education is and a liberal government cannot simply presume something that is in dispute.¹⁰³ It is unrealistic to expect that a consensus on teaching approaches can be achieved even in a well-ordered society. Such a consensus may come to exist among strongly just citizens, but not among weakly just citizens.¹⁰⁴

Hence, if an overlapping consensus depends on the production of a particular kind of citizen, and this production depends on a consensus on teaching approaches, and no particular teaching approach can be generally accepted in the well-ordered society, then education is not a good answer to the problem of general acceptability of Rawls' contractarianism. It is because the educational approach itself also cannot avoid controversy even in the well-ordered society.

¹⁰³ Kukathas (2001: 325).

¹⁰⁴ Some might suggest that, in order to secure the realization of the two moral capacities of children, a government should not permit these weakly just citizens to teach their children something that is incompatible with the political conceptions. The government should ensure that all children receive proper education even by infringing citizens' rights. However, citizens' rights of teaching children are closely related to basic liberties such as the liberty of conscience, and these basic liberties are necessary for the development of the two moral capacities. If these citizens' rights of teaching are infringed, their development of two moral capacities of these citizens will also be harmed. Hence, although this educational policy protects the two moral capacities of children, it harms the two moral capacities of the weakly just citizens at the same time. Also, I am not sure whether Rawls would agree with this educational policy because he prefers a looser approach of education. For an example of this suggestion, see Quong (2011: 301-305).

Moreover, if the general acceptability of political conceptions presupposes so many educational procedures, then this general acceptability is still vulnerable to the problem of immigration. Rawls presupposes that the just society is ‘a closed society: that is, we are to regard it as self-contained and as having no relations with other societies’.¹⁰⁵ Yet this assumption is unrealistic since immigration is inevitable in the modern world. Many people cross national boundaries and move to different countries. We can expect that there would be a large number of people trying to migrate to the well-ordered society that Rawls describes. These immigrants would not have received citizenship education before, so they would not be the strongly just citizens that Rawls expects. They may hold other unreasonable comprehensive doctrines and may not be willing to acknowledge the special status of political conceptions. Because of this lack of learning experience, the assumptions of justice as fairness would not be generally acceptable to them. They would be outside the scope of overlapping consensus. However, if the well-ordered society does not allow these immigrants to move in and sets a very high standard of immigration, then its migration policy would be unrealistic. This further shows the limit of hybrid contractarianism; if hybrid contractarianism can be applied only in a closed society which implements an extensive educational policy, then nearly no existing societies can fit this standard.

In conclusion, Rawls overestimates the educational effect of the social circumstances of the well-ordered society. Undoubtedly the well-ordered society is a desirable environment in that it cultivates children to become strongly just citizens. However, even in such an ideal environment, the number of people who will become strongly just citizens is still limited.

¹⁰⁵ Rawls (1993: 12).

7.5 The desirability of justice as fairness

If the social circumstances cannot provide a satisfactory answer to the question of general acceptability, then Rawls might have to rely on philosophical arguments to explain why citizens should transform their comprehensive doctrines. As we saw in section 2, comprehensive doctrines are not fully comprehensive and rigid. They usually have some conceptual spaces and citizens can develop an independent allegiance to support political conceptions. But why would they develop this allegiance? Rawls explains that this allegiance can be attributed to their recognition of the benefits brought by justice as fairness. Surely Rawls does not overlook the possibility of the conflicts between political conceptions and other elements in the comprehensive doctrines. But he believes that, in light of the great desirability of justice as fairness, citizens in the well-ordered society would become strongly just citizens and would transform their doctrines into reasonable comprehensive doctrines after reflection. When they conflict with each other, political conceptions ‘normally outweigh whatever other values oppose them’.¹⁰⁶ Rawls gives two answers to explain the desirability of a well-ordered society.¹⁰⁷ This section considers whether these two reasons are enough to persuade citizens to transform their doctrines.

¹⁰⁶ Rawls (1993: 209).

¹⁰⁷ Actually Rawls gives three reasons, and the third reason is that following principles of justice is the most effective way to exercise two moral capacities. Because of the Aristotelian Principle, people usually have ‘higher-order interests’ in exercising these two moral capacities. See Rawls (1993: 202-203). However, the problem of this argument has been discussed in the last chapter already. So I will only focus on the other two reasons that Rawls provides.

7.5.1 The collective-social-good argument

The first reason relies on the collective social good of the well-ordered society. This argument first appears in *A Theory of Justice*,¹⁰⁸ and Rawls continues to use it in *Political Liberalism*.¹⁰⁹ Rawls first emphasizes the social nature of mankind. The power of a person is limited: ‘no one person can do everything he might do; nor a fortiori can he do everything that any other person can do. The potentialities of each individual are greater than those he can hope to realize.’¹¹⁰ Each person can only realize in himself one of the many possible lives that together make up human flourishing. This implies that if people choose to develop some of their talents, then they inevitably fail to enjoy the benefits which come from the development of other talents. But if people participate in social cooperation, they can benefit from the fruition of others’ developments. In a well-ordered society, principles of justice guarantee every member of society equal basic liberties and adequate primary goods. All members are given sufficient freedom and resources to realize their potentialities and participate in the total sum of the realized natural assets of the others. These members together form a rich and diversified collective activity which would benefit every participant. Each citizen enjoys this social good ‘through citizens’ joint activity in mutual dependence on the appropriate actions being taken by others’.¹¹¹ Rawls believes that this kind of social good which is based on cooperation should not be unfamiliar to people.

That there should be such a political and social good is no more mysterious than that members of an orchestra, or players on a team, or even both teams in a game, should take pleasure and a certain (proper) pride in a good performance, or in a good play of the game, one that they will

¹⁰⁸ Rawls (1999a: 463).

¹⁰⁹ Rawls (1993: 204).

¹¹⁰ Rawls (1999a: 458).

¹¹¹ Rawls (1993: 204).

want to remember. No doubt the requisite conditions become more difficult to satisfy as societies become larger and the social distance between citizens becomes greater, but these differences, as great and inhibiting as they may be, do not affect the psychological principle involved in realizing the good of justice in a well-ordered society.¹¹²

A well-ordered society is a community of humankind in which its members are able to enjoy one another's excellence and individuality. Given that no one can realize their conception of the good on their own without relying on others, everyone can benefit from this collective asset. That is why Rawls argues the well-ordered society is 'a great social good and appreciated as such'.¹¹³

However, this possible response overemphasizes the dependence of comprehensive doctrines on the well-ordered society. Suppose that some weakly just citizens belong to one group. If the size of this group is large enough, then they are able to sustain themselves. They can separate themselves from the society and refuse to give political conceptions special status. For example, as long as a religious community can sustain itself, why do its members have to accept the rule of a liberal state which may prohibit the realization of their religious goods? The ideal cooperative circumstance presented by Rawls can also be achieved by a group of citizens given that this group is large enough. Surely, one may argue that, through participating in social cooperation with other members, these particular associations can obtain more goods and enjoy the excellences of other members. Nevertheless, the benefit which comes from social cooperation may still fail to overcome the loss which they have to suffer when they make a concession to justice as fairness. Isolating themselves from the society may still be a better choice for them to realize their religious doctrines.

¹¹² Rawls (1993: 204).

¹¹³ Rawls (2001: 201).

Even if we believe that the social good of a well-ordered society is a necessary condition of realizing the comprehensive conceptions of the good, it does not mean that everyone should act as free and equal citizens *unconditionally*. It seems that a more intelligent strategy is to act as free and equal citizens *conditionally*. Obviously, sometimes breaking principles of justice could help citizens to realize their conceptions of the good more effectively. On the other hand, we know that the whole social cooperative scheme will not vanish because of one or two rule-breaking behaviours. Therefore, instead of being a loyal follower of principles of justice, one should be a ‘sensible knave’, that is, acting as a free and equal citizen ordinarily, but waiting for chances to break the rules with impunity. In this way one can enjoy the fruit of social cooperation in the well-ordered society and also take advantage of others by breaking the rule as long as this will not affect one’s ability to remain with the well-ordered society. The fact that one can benefit from the citizens’ joint activity in mutual dependence can at most show that one should act upon principles of justice conditionally. It cannot explain why citizens should not insist on holding reasonable comprehensive doctrines if they can do it with impunity and still enjoy the collective good of a well-ordered society. In the end, although Rawls is correct that a well-ordered society governed by justice as fairness is a ‘highly significant’ social good,¹¹⁴ he exaggerates the importance of this social good on the realization of comprehensive conceptions of the good.

7.5.2 The self-respect argument

The second reason that Rawls could give is that justice as fairness ‘secure[s] for citizens...the social bases of mutual respect’.¹¹⁵ Rawls believe that self-respect, which is related to the exercise of two moral capacities, is ‘perhaps the most

¹¹⁴ Rawls (2001: 201).

¹¹⁵ Rawls (1993: 203).

important primary good'.¹¹⁶ If people can exercise their two moral capacities adequately, then they achieve a moral ideal (free and equal citizenship) that is integral to their self-conceptions and may enjoy a certain kind of moral excellence. This achievement is important because it 'provides a secure sense of our own value, a firm conviction that our determinate conception of the good is worth carrying out. Without self-respect nothing may seem worth doing, and if some things have value for us, we lack the will to pursue them'.¹¹⁷ If people fail to achieve this ideal, then they will suffer from a sense of shame, that is, 'an injury to self-respect'.¹¹⁸ In order to protect self-respect and avoid shame, citizens will support justice as fairness and transform their doctrine into reasonable comprehensive doctrines. This is because justice as fairness can provide equal basic liberties and an adequate quantity of primary goods, which are essential for the development of two moral capacities. Given that these primary goods are guaranteed, the self-respect of citizens can be protected. 'Only the two principles of justice guarantee the basic liberties, they are more effective than the other alternatives in encouraging and supporting the self-respect of citizens as equal persons'.¹¹⁹ Hence, by living in a well-ordered society, each citizen can secure their fundamental need for self-respect and avoid suffering from shame.

This reason explains the desirability of justice as fairness by the sense of worth that it protects. However, although a just institution can help citizens to secure self-respect, it can at most provide *one of the ways* of securing self-respect. Self-respect can be established in many alternative ways. Rawls defines self-respect in quite broadly way. He explains self-respect in terms of the achievement of the goals and ideals one

¹¹⁶ Rawls (1999a: 386).

¹¹⁷ Rawls (1993: 318).

¹¹⁸ Rawls (1999a: 388).

¹¹⁹ Rawls (1993: 319).

incorporates into one's life plans.¹²⁰ Rawls distinguishes two ways that self-respect can be harmed. First is the case of moral shame. Moral shame refers specifically to the failure to satisfy the desire to realize the ideal of free and equal citizen. As we saw before, although individuals treat the conception of free and equal citizen in different ways, all of them accept this ideal and take the realization of this ideal as an excellence. Hence, if citizens fail to realize this ideal, then it follows that they fail to do something that they 'prize as excellences',¹²¹ and fail to achieve an excellence that is publicly recognized by other fellow citizens. This creates in them a sense of moral shame and harms their self-respect. Apart from this, there is also another kind of shame - natural shame. Natural shame refers to the failure to satisfy 'our aspirations, what we try to do and with whom we wish to associate'.¹²² Its source is much broader than moral shame because it can come from the failure of satisfying any goal and ideal, provided that these goals and ideals play important roles in a plan of life. An athlete can be ashamed if he fails to win a championship. A student can be ashamed if he fails to become the kind of exemplary student that he expects to be. 'Given our plan of life, we tend to be ashamed of those defects in our person and failures in our actions that indicate a loss or absence of the excellences essential to our carrying out our more important associative aims'.¹²³

Given that there are two kinds of shame, embracing principle of justice might help citizens avoid moral shame, but it might cause citizens to suffer from natural shame. Citizens can avoid moral shame because they can become the free and equal citizens they want to be. But apart from this ideal, citizens also have many other ideals and goals in their comprehensive doctrines. These ideals and goals may conflict with justice as fairness. An athlete may want more primary goods, so that he can have

¹²⁰ Rawls (1999a: 386-387).

¹²¹ Rawls (1999a: 390).

¹²² Rawls (1999a: 390).

¹²³ Rawls (1999a: 390).

more resources to support his training. A religious believer may want to change political arrangements for the sake of attracting more people to believe in her religion. Failure to achieve these goals and realize these ideals may create a strong sense of natural shame. Moreover, it is hard to see that natural shame is any less painful than moral shame. Rawls does not explain why people would choose to avoid moral shame but not natural shame. If shame is a reason for citizens to support justice as fairness, then it seems it can be a reason for citizens to go against justice as fairness as well. Citizens who choose to support a just institution may have to suffer from natural shame even though they avoid moral shame. Their self-respect might still be harmed.

Furthermore, this interpretation of self-respect is also controversial in the well-ordered society. As John Deigh points out, Rawls' account of self-respect is inadequate because it fails to explain certain expression of self-respect and shame.¹²⁴ Sometimes there is a loss of self-respect but we do not feel any shame (imagine a tennis player who tries his best but is still defeated by a much stronger opponent). Whether or not Deigh's critique is sound, it seems too quick to assume that Rawls' interpretation of self-respect would be uncontroversial in a well-ordered society where every citizen can freely exercise their human reason. Given that citizens have different philosophical and religious doctrines, they should also have different understandings of self-respect. If Rawls simply assumes that his interpretation of self-respect is generally accepted in a well-ordered society, then it merely begs the question.

In conclusion, these two explanations use the same strategy: a strategy of emphasizing the desirability of adopting a reasonable comprehensive doctrine and

¹²⁴ Deigh (1983: 243-245).

living in a well-ordered society. The first explanation argues the desirability from a social perspective, while the second argues it from an individual perspective. However, this strategy is inadequate since it can at most show that justice as fairness is desirable, but it is not enough to explain why citizens should change their doctrines into reasonable comprehensive doctrines. Why *must* they be strongly just citizens? The desirability of justice as fairness is not irreplaceable. Citizens can enjoy other more desirable goods even though they are weakly just citizens. Therefore, although Rawls gives some reasons to explain why citizens should join his social contract, this is not enough to show that citizens would generally accept them even after reflection.

7.6 Conclusion

At the end of the last chapter we saw that Rawls faces a dilemma in choosing to meet the demands of generality and priority. Either he has to keep the Kantian interpretation of human nature and sacrifice generality, or he has to drop the Kantian interpretation and sacrifice priority. Thus the later Rawls chooses to drop the Kantian interpretation but suggest a new, more ‘political’ version of contractarianism which can satisfy both the condition of generality and the condition of priority. He first secures the condition of priority by defining several freestanding political conceptions implicit in the public political culture. These political conceptions represent a way of engaging in rational and reasonable thinking in the political domain. If rationality and reasonableness are understood in this particular context, then they would lead to the same set of principles: justice as fairness. By showing that these principles are both rationally and reasonably justified, Rawls’ hybrid contractarianism can satisfy the condition of priority. Rawls believes that this approach to satisfying the condition of priority can also avoid the problem of generality that troubles the earlier version of his contractarianism. He first acknowledges that his contractarianism can only be applied to liberal democratic societies. But he believes that, within these societies, the political conceptions are

generally acceptable. Thus the conception of justice as fairness justified by these political conceptions is an overlapping consensus. Hence, by making a 'political' turn, Rawls' hybrid contractarianism can satisfy the condition of priority without sacrificing the condition of generality.

However, I believe that the dilemma is more serious than Rawls thinks. Even in the most idealist liberal democratic society that Rawls expects, that is, a well-ordered society governed by justice as fairness, political conceptions are still not generally accepted by citizens. In a well-ordered society, citizens have opportunities to learn political conceptions and undergo the three-stage moral development described by Rawls. However, it only means that political conceptions could be built into citizens' comprehensive doctrines and they would become weakly just citizens. It is too much to assume that all citizens would become strongly just citizens who are willing to take these political conceptions as the *sole* ground of justification of political principles. Because of the free exercise of human reason, citizens are able to develop different comprehensive doctrines. Although political conceptions are widely shared among citizens, these conceptions are only *part* of their comprehensive doctrines. Why would this part, which is constituted by merely freestanding intuitive ideas, have any special status? If its special status is justified by political conceptions themselves, then this undermines the freestanding nature of political conceptions. If its special status is justified by other parts of comprehensive doctrines, then a lot of comprehensive doctrines do not acknowledge this special status. In the end, Rawls still fails to avoid the dilemma of having to choose between generality and priority. If his theory has to satisfy the condition of priority, then he must presuppose that every citizen gives a particular 'political' way of engaging in practical reasoning an overriding importance, yet this presupposition cannot be generally accepted even in the well-ordered society that Rawls expects.

The failure of the later Rawls indicates how difficult it is to satisfy both the condition of generality and the condition of priority. Rationality and reasonableness are the two aspects of practical reason, but they justify different conclusions when they are applied in different contexts. An artist can apply rationality and reasonableness in his own way, while a communist can apply them in another way. Therefore, when a contractarian wants to use practical reason as the source of the normative power of his contractarian theory, he can only define practical reason in a very formal way, for the sake of being compatible with different applications of practical reason. For example, Gauthier simply defines rationality as formal principles of rational choice. No matter what ends a person has, given that he applies this instrumental conception of rationality, he is subjected to the normativity of Hobbesian contractarianism. He makes this assumption to ensure that their assumptions are generally acceptable. However, since Rawls has to show that rationality and reasonableness can be reconciled with one another, he assumes a particular way of engaging in rational and reasonable thinking, which is a way used by 'free and equal citizens'. Nevertheless, Rawls eventually fails to explain why it is generally acceptable that this particular way might enjoy a special status but other ways may not.¹²⁵ While Rawls preserves the virtue of priority, he inevitably gives up the virtue of generality.

¹²⁵ Kenneth Baynes made a similar point that Rawls only describes the conception of free and equal person as *one* of the possible account of our self-conceptions, but fails to show why it is the best account that we must take. See Baynes (1992: 30).

Chapter 8 Conclusion

In *Elements of the Philosophy of Right*, Hegel famously said, ‘the Owl of Minerva begins its flight only with the onset of dusk’.¹ This dictum can also be applied to the history of thought because a form of thought can only be fairly evaluated once developed maturely. Like the seventeenth and eighteenth centuries, the late twentieth century is another golden age of contractarianism and contractarians have employed this methodology in various ways since the 1960s. Today is nearly half a century after the rebirth of contractarianism and is an appropriate time to evaluate the achievement of this philosophical movement. The aim of this thesis is to be the Owl of Minerva which spreads its wings during the twilight of contemporary contractarianism. Now the flight is reaching its end. Before the end of this flight I want to discuss three things. First, I will briefly outline the argument of this thesis and explain the main contribution. Secondly, I will show why my analysis is better than, and can contribute to, existing approaches to evaluating contractarianism. Lastly, I will briefly discuss how, if my analysis is true, contractarianism might be developed in the future.

8.1 The dilemma of contractarianism

This thesis questions whether or not contractarianism is a good methodology for justifying political principles. By looking at different contemporary contractarian

¹ Hegel (1991: 23). In fact the original meaning of this sentence is that we can only grasp the rationality in history and politics by philosophy. Major historical transformations in political and social institutions were usually realized even though people did not fully understand the meanings of these transformations. Hence philosophers have to look back the history and evaluate these historical events after these events have been over. My usage here is a bit different from Hegel’s original usage because the objects of evaluation are theories in the history of political thought rather than political and social events in the history.

Chapter 8 Conclusion

models, it is possible to see the strengths and weaknesses of contractarianism as a general methodology. Unlike classical contractarianism, contemporary contractarianism discards voluntarism and takes practical reason as the source of normativity. Based on different conceptions of practical reason, different contractarian models can be developed. However, not all models are the same good. Some can provide more convincing justifications, while some cannot. This depends on whether the model can satisfy two conditions: generality and priority. This thesis discusses different models and the result of my analysis can be presented in the following table:

Table 5 Contractarianism's Dilemma

	Satisfies the condition of generality	Does not satisfy the condition of generality
Satisfies the condition of priority		Hybrid contractarianism (Rawls)
Does not satisfy the condition of priority	Hobbesian contractarianism (Gauthier), Kantian contractarianism (Scanlon, Barry)	Utilitarian contractarianism (Harsanyi)

Chapter 8 Conclusion

The first possible contractarian model is Utilitarian contractarianism, which is represented by Harsanyi and attempts to justify utilitarianism by a contractarian model. This model is based on an altruistic conception of rationality. However, since this conception is not a general pattern of practical reasoning, utilitarian contractarianism fails to be generally acceptable. Also, this conception is not inclusive because it ignores the conception of reasonableness, which is the other aspect of practical reason. Hence this model fails to satisfy either of the conditions of generality and priority.

The second model is Hobbesian contractarianism which is indebted to Hobbes and represented by Gauthier. This model is based on an instrumental conception of rationality. Since this conception is merely a formal account, it is generally compatible with rational reasoning, but its flaw is that it overlooks another dimension of practical reason: reasonableness.

Kantian contractarianism, the third model, also has a similar theoretical problem. It is represented by Scanlon and has an historical root that can be traced back to Locke and Kant. Contrary to Hobbesian contractarianism, its conception of practical reason is a conception of reasonableness that takes the constraint of justifiability as absolute. Our sense of guilt proves that this conception grasps a general feature of our practical reasoning. Nevertheless, it is one-sided and overlooks another aspect of practical reason: rationality. Hence, both Hobbesian and Kantian contractarianism share the same problem: they can satisfy the condition of generality, but fail to satisfy the condition of priority.

The problem of these two models can be overcome in the fourth model: hybrid contractarianism. Like Rousseau's, Rawls' contractarian model has a hybrid character. Rawls' conception of practical reason is a dual conception which consists of both rationality and reasonableness. Therefore his social contract is both rationally

and reasonably justified. By embracing this social contract, rationality and reasonableness are congruent with each other. This model can satisfy the condition of priority because of its more comprehensive conception of practical reason. Nevertheless, since Rawls has to prove the congruence between rationality and reasonableness, he relies on some substantial conceptions of practical reason. While the earlier Rawls relied upon a conception of rationality incorporating the Kantian interpretation of human nature, the later Rawls assumes a political approach to engage in rational and reasonable thinking. These substantial conceptions of practical reason are not generally acceptable to people, even in the liberal democratic society to which the later Rawls restricts the application of his theory. Hence, hybrid contractarianism satisfies the condition of priority at the cost of losing generality.

Through studying various contractarian models we can uncover the inherent dilemma of contemporary contractarianism, which is also the main contribution of this thesis. Although various contractarian models were developed, these models can satisfy either the condition of generality or the condition of priority. Insofar as contemporary contractarianism assumes practical reason as the source of normativity, this dilemma is unavoidable. This is because no conception of practical reason can develop a contractarian theory which can satisfy both the conditions of generality and priority. Practical reason has two aspects: rationality and reasonableness. The former is related to the idea of *good*, while the latter is related to the idea of *right*. The content of a conception of practical reason is largely determined by how it defines the relationship between these two aspects. Some contractarians are only concerned with one aspect. This aspect is the whole conception of practical reason and another aspect is only a part of this aspect. Yet political principles justified by only one of the two aspects cannot assume a supreme authority. Some contractarians understand these two aspects as being independent of one another and argue that they justify the same set of political principles. Yet these contractarians underestimate the tension between rationality and reasonableness within practical reason. Rationality and reasonableness always lead to different conclusions; it is usually rational to be

unreasonable as well as reasonable to be irrational. If contractarians want to argue for the congruence between rationality and reasonableness, then they cannot avoid assuming a controversial conception of rationality or reasonableness. Thus, no matter how contractarians construct their theory, they still have to face the dilemma to choose between generality and priority.

Because of this dilemma, contractarianism can at most justify principles which are general but not overriding, or principles which are overriding but not general. Since a political principle should be both general and overriding, contractarianism is bound to fail to provide a satisfactory justification for political principles. This also implies that the problems of Hobbesian, Kantian and hybrid contractarianism are due not to the poor application of the methodology, but, rather, to the dilemma inherent in this methodology. Once contractarians rely on this methodology to justify political principles, they are destined to be caught between the demands of generality and the demand of priority.²

8.2 Contributions to the existing approaches

This thesis does not only show the dilemma inherent in contractarianism, but also contributes to the existing discussions about this methodology by moving it to a deeper level. In Chapter 1, I mentioned that there are usually three approaches to evaluating this methodology. First, people may simply praise the strength of this

² However, one should not extend my argument to a critique to contractarian theories which aims at justifying moral principles. As I mentioned in Chapter 1, the nature of political principle is different from that of moral principle. For example, moral principles need not be both general and overriding. Some moral theorists argue that ‘overridingness’ is not a necessary feature of morality (Scheffler 1992: 56). So moral principles may only need to be general. Contractarianism may be a good methodology for justifying moral principles, but this function of contractarianism is not the object of my discussion. The purpose of my thesis is only to examine whether contractarianism can provide a good justification for *political* principles.

methodology.³ Secondly, people may criticize contractarianism as irrelevant because of the dubious relationship between the hypothetical contract and the real world.⁴ Thirdly, people may only discuss and criticise particular contractarians.⁵

My general discussion of contractarianism can provide a more comprehensive picture than any of these approaches and can also show what is inadequate in these approaches. With respect to the first approach, the three virtues of contractarianism that I laid out in Chapter 2 can explain why so many philosophers were attracted by this tradition. Nevertheless, contractarianism also has to face the dilemma to choose between generality and priority. This illustrates the limitations of this methodology. It is not as desirable as some philosophers expect.

The second approach is a standard critique of contractarianism that a hypothetical contract cannot bind people since the source of normativity of contractarianism is unclear. This critique has the merit of pointing out a distinction between hypothetical, and actual consent. Although this distinction is correct, it is too quick to say that a hypothetical contract is irrelevant to the real world. Sympathetic critics acknowledge that some hypothetical contracts, such as Rawls' and Gauthier's contract, have normative force, but when they track the source of normativity of contractarianism, the results are either mistaken or vague. Some argue that contractarianism presupposes natural rights.⁶ Some find the source of normativity in the 'rational will' but do not go into detail.⁷ These critics rightly point out that hypothetical contract is a heuristic device, but they fail to indicate what it is heuristic *for*. In this thesis, I showed that the source of normativity of contractarianism is

³ Gauthier (1991a)

⁴ Dworkin (1973), Brudney (1991) and Hampton (1993b).

⁵ Boucher and Kelly (1994a) and Vallentyne (1991a).

⁶ Dworkin (1973: 51-52).

⁷ Stark (2000: 333-334)

practical reason, which consists of rationality and reasonableness. Although practical reason has an unconditional and overriding normative force, people are unclear about what political principles can be justified by practical reason. Thus they have to rely on the hypothetical contract, which is a heuristic device used for deriving substantial political principles from a conception of practical reason. The hypothetical contractors represent how actual people would think if they could engage in practical reasoning correctly. By identifying themselves with hypothetical contractors, actual people can *discover* the correct political principles that they should legislate for themselves. The basic strategy of contractarianism is to make use of practical reason to explain why certain political principles have an unconditional and overriding normative force.

My general analysis of contractarianism also discovers the real significances of critiques of particular contractarians in the third approach. For example, there are a number of critics who point out that Hobbesian contractarianism goes against some of our moral intuitions.⁸ Although these critiques are sound, they cannot constitute a strong criticism of Hobbesian contractarianism. Some Hobbesian contractarians, such as Gauthier, would simply bite the bullet and acknowledge that their contractarian principles go against some moral intuitions. However, from their perspective, what should be rejected are not their contractarian principles but rather these intuitions, since these intuitions are irrational, and thus unreliable. Rationality has a more crucial role in practical reasoning than these fragmentary intuitions do.⁹ The strength of these critiques can be made more explicit if they are understood in the broader context provided in this thesis. As we have seen, these moral intuitions are not just fragmentary intuitions. Rather, they originate in reasonableness, which is another aspect of practical reason. Given that rationality and reasonableness are

⁸ Hampton (1991: 48-49) and Barry (1995a: 42).

⁹ Gauthier (1986: 269).

equally important in practical reason, Hobbesian contractarians lose the ground for claiming that those moral intuitions are more dispensable. From this example we can see how this thesis can shed light on the existing critiques of particular contractarians. These critiques have never been scarce, but their real strength sometimes can only be seen when they are placed in a bigger picture.

Through examining the inadequacies of these three existing approaches we can also see why a general analysis of contractarianism is necessary. All three approaches make a strong contribution to the critical study of contractarianism. Nevertheless, without a general analysis on contractarianism, these approaches are incomplete and their real significance remains obscure. The first approach ignores the weakness of this methodology. The second approach recognises the problem of the source of normativity, but does not go deeply enough into this problem. The third approach only focuses on particular contractarians and lacks a broad vision. The analyses of these three approaches can be further improved when they are understood in light of a more comprehensive and deeper analysis on contractarianism. Hence, a general analysis of contractarianism is not only necessary for evaluating the worth of contractarianism itself, but also necessary for evaluating the existing critiques of this methodology.

8.3 The future of contractarianism?

This thesis uncovers the dilemma of contractarianism, so a number of questions naturally arise: How can contractarians develop after recognizing this dilemma? What will the future of this school be? Contractarians could give up the ambition to justify general and overriding political principles and sacrifice either of these properties. However, as I argued in Chapter 1, both generality and priority are the necessary properties of political principles. Given that both of these two properties are necessary, but contractarianism cannot justify principles which have both

properties, does it indicate the end of the era of contractarianism in political philosophy?

Contractarians who want to rescue this methodology might defend it by finding an *objective* source of normativity instead of practical reason. Justifying principles by practical reason belongs to the *intersubjective* approach. The normativity of principle is based on the capacities for practical reason shared among subjects. Provided that people are rational and reasonable, they are obligated to this principle. However, as we have seen, a principle justified by this approach cannot avoid the dilemma that arises when it becomes necessary to choose between generality and priority, since the tension between rationality and reasonableness is a common feature of practical reason. This implies that if contractarianism has to rid itself of this dilemma, it has to adopt another approach to explain the source of normativity.

Suppose that contractarians turn to justify their social contract by an objective source of normativity. They describe the social contract as representing an ideal form of society that actual people *ought* to appreciate, no matter what actual people rationally or reasonably think. The normativity of this social contract is objective in the sense that people are obligated to it independent of subjective will. By arguing that the normativity is objective, this approach should be able to escape from the dilemma to choose between generality and priority. In this objective approach, contractarians could directly require that the social contract should be generally acceptable and overriding. These two properties of the social contract are justified by an objective moral authority (or an objectively desirable moral ideal) rather than being justified by the subjective will of people. Hence, if the social contract is not generally acceptable or overriding in the real world, then it is only because of the flaws and weaknesses of actual people. The general acceptability and the overridingness of the social contract are unaffected. Indeed, this inclination of ‘objectivization’ can also be seen in the latter writings of some contemporary contractarians. In *Moral*

Dimensions, Scanlon argues that the duty to his contractualist principles is independent of what *ends* people have and how they think about these principles.¹⁰ Even if people disagree with these principles after careful deliberation, they still *ought* to follow these contractualist principles. It is unclear whether Scanlon makes this change because he recognizes the dilemma of contractarianism, but this dilemma can be avoided by adopting an objective approach to justification.

Although the ‘objectivisation’ move can possibly rescue contractarianism from the dilemma, contractarianism will further lose the virtues that it originally possessed, especially the virtue of naturalism. As noted in Chapter 2, one of the reasons that contractarianism is so appealing to philosophers is that it is compatible with the secular, scientific worldview, which is prevalent in the modern world. Since it takes subjective capacities as the source of normativity, it does not need to presuppose any external, objective normative authority to justify principles. To contractarianism, a normative principle is something which is *agreed* among subjects in certain appropriate circumstances, but not something objectively existing ‘out-there’, waiting for people to recognize, and binding people unconditionally. This is an important virtue because, in a scientific worldview, it is difficult to prove the existence of an external normative authority and explain its relationship with other natural facts. However, if contractarianism made the ‘objectivization’ move, then its source of normativity would rely on an external, objective normative authority. Only when this external normative authority was assumed, could contractarians argue that the normativity of social contract is independent of the subjective capacities of people. But then contractarians would have to suffer from the cost of losing the virtue of naturalism. They would have to explain the metaphysical status of this external authority, how it is related to the natural world, and why it has normative force. Although this ‘objectivisation’ move can help contractarianism avoid facing

¹⁰ Scanlon (2008: 99-100).

the dilemma to choose between generality and priority, it leads to another serious problem.

Moreover, if objectivisation is really the future direction of contractarianism, then it seems ironic, for in this process contractarianism would only be borrowing the cloak of the social contract to drape around a natural law theory. Natural law theorists, such as Thomas Aquinas, adopt an objective approach to explaining the source of normativity and argue that normative principles are objective, eternal laws. They justify the normativity of these principles by a kind of obligating authority external to subjects, such as God, for example. At the beginning, contractarianism originated as a reaction to the natural law school, which was dominant until the sixteenth century.¹¹ Its dominant role was replaced by contractarianism for different reasons. One of these reasons was that, due to the emergence of modern science, the order of the world was more and more understood as mechanistic and naturalistic. External, independent natural laws seemed to be *odd* in this worldview. Hence the era of the natural law school ended, and contractarianism rose in the seventeenth and eighteenth century. However, if the future of contractarianism is to rely on an objective, external source of normativity, then it would mean returning to the natural law school. At first, contractarians, dissatisfied with the natural law school, took voluntary consent to be the source of normativity. Then voluntarism was discarded and contractarians turned to explain the source of normativity in practical reason. But if practical reason is an unsatisfactory source as well, then at last contractarians may

¹¹ However, some contractarians also assume the existence of natural law in their contract theories. For example, although Hobbes and Locke were the representatives of classical contractarians, they also adopted the assumption of natural law. Nevertheless, these contractarians are still clearly different from the medieval natural law theorists. For natural law theorists, it is assumed that there is something objectively good, which determines the natural laws and what a good life is. But Hobbes and Locke disagreed with these themes. Their assumption of natural laws merely means there is something that people 'should' do. Also, the fact that Hobbes and Locke were fiercely criticised by the natural law theorists at that time, can also show the difference between these classical contractarians and the natural law theorists. I am indebted to Chandran Kukathas for these observations.

Chapter 8 Conclusion

have to seek an external source of normativity, like the natural law school. Contractarianism originates from a departure from the natural law tradition, but may end by returning to the natural law tradition. The Owl of Minerva ends its flight, only to find that it has returned to the tree from which it set out.

Bibliography

- Allais, Maurice (1979). 'The So-called Allais Paradox and Rational Decision under Certainty', in Maurice Allais and Ole Hagen (ed.), *Expected Utility Hypothesis and the Allais Paradox* (Dordrecht, Netherland: Reidel Publishing).
- Anscombe, G.E.M. (1981). 'On the Source of the Authority of the State', in G. E. M. Anscombe, *Ethics, Religion and Politics: Collected Papers, vol. 3* (Minneapolis: University of Minnesota Press), pp. 136-154
- Aquinas, St. Thomas (2002). *Political Writings* (Cambridge, U.K.: Cambridge University Press)
- Arrow, Kenneth (1963). *Social Choice and Individual Value* (New York: Wiley).
- (1967). 'Public and Private Values,' in Sydney Hook (ed.), *Human Values and Economic Policy* (New York: New York University Press), pp. 2-21.
- Ashford, Elizabeth (2003). 'The Demandingness of Scanlon's Contractualism,' *Ethics* 113(2); 273-302.
- Atiyah, P. S. (1979). *The Rise and Fall of Freedom of Contract* (New York: Oxford University Press).
- Audard, Catherine (2007). *John Rawls* (Stockfield: Acumen Publishing)
- Baier, Annette C. (1994). 'What do Women Want in a Moral Theory?' in her *Moral Prejudice* (Cambridge, MA.: Harvard University Press)
- Baier, Kurt (1958). *The Moral Point of View* (Ithaca and London: Cornell University Press)
- (1988). 'Rationality, Value and Preference' in Paul, Miller, Paul and Ahrens (1988), pp. 17-45.
- (1989). 'Justice and the aims of Political Philosophy', *Ethics*, 99(4): 771-790.
- Barker, Sir Ernest (1960). *Greek Political Theory: Plato and his predecessors* (London: Methuen)
- (1980). *Social Contract: Essays by Locke, Hume and Rousseau*. (Westport, Conn.: Greenwood Press, 1980).
- Baron, Marcia (1991). 'Impartiality and Friendship,' *Ethics*, 101(4): 836-857.
- Barry, Brian (1989). *Theories of Justice* (Berkeley and Los Angeles, California: University of California Press)
- (1995a). *Justice as Impartiality*. (Oxford: Clarendon Press).
- (1995b). 'John Rawls and the Search for Stability', *Ethics*, 105(4): 874-915
- (2000) 'Something in the Disputation not Unpleasant,' in Paul Kelly (ed.), *Impartiality, Neutrality and Justice: Re-Reading Brian Barry's Justice as Impartiality* (Edinburgh University Press, 2000), pp. 186-257.
- Bates, Stanley (1974). 'The Motivation to be Just', *Ethics*, 85(1): 1-17
- Baynes, Kenneth (1992). 'Constructivism and Practical Reason in Rawls', *Analyse & Kritik*, 14: 18-32.
- Becker Lawrence C., and Charlotte B. Becker (ed.) (2001), *Encyclopedia of Ethics*, 2nd edition (New York: Routledge)
- (2005). 'Reciprocity, Justice, and Disability,' *Ethics*, 116(1): 9-39

Bibliography

- Bicchieri, Cristina (2004). 'Rationality and Game Theory' in Alfred Mele and Piers Rawling (ed.), *The Oxford Handbook of Rationality* (New York: Oxford University Press), P. 182-205.
- Binmore, Ken (1994). *Game Theory and the Social Contract Vol. 1: Playing Fair* (Cambridge, MA.: M.I.T. Press)
- (2005). *Natural Justice* (New York: Oxford University Press)
- (2008). 'Naturalizing Harsanyi and Rawls' in Marc Fleurbaey, Maurice Salles and John A. Weymark (ed.), *Justice, political liberalism, and utilitarianism : themes from Harsanyi and Rawls* (New York: Cambridge University Press), pp. 303-333.
- Bohman, James (1995). 'Public Reason and Cultural Pluralism: Political Liberalism and the Problem of Moral Conflict', *Political Theory* , 23(2): 253-279
- Boucher, David and Paul Kelly (1994a). *The Social Contract From Hobbes to Rawls* (London: Routledge).
- (1994b). 'The social contract and its critics: an overview', in Boucher and Kelly (1994a), pp. 1-34.
- Brandom, Robert (2001). 'What do Expression of Preference Express?', in Morris and Ripstein (2001), pp. 11-36.
- Brink, David O. (1987). 'Rawlsian Constructivism in Moral Theory' *Canadian Journal of Philosophy*, 17(1): 71-90
- (1997). 'Kantian Rationalism: Inescapability, Authority, and Supremacy', in Cullity and Gaut (1997), pp. 255-292.
- Broome, John (1991). *Weighing goods: equality, uncertainty and time* (Cambridge, MA.: Basil Blackwell)
- Brower, Bruce (1994). 'The Limits of Public Reason', *The Journal of Philosophy*, 91(1): 5-26
- Brown, Henry Phelps (1988). *Egalitarianism and the Generation of Inequality* (Oxford: Clarendon Press)
- Brudney, Daniel (1991). 'Hypothetical Consent and Moral Force', *Law and Philosophy*, 10(3), pp. 235-270.
- Buchanan, Allen (1975). 'Revisability and Rational Choice', *Canadian Journal of Philosophy*, 5(3): 395-408
- (1990). 'Justice as Reciprocity versus Subject-Centered Justice,' *Philosophy and Public Affairs*, 19(3): 227-252
- (2000). 'Justice, Legitimacy, and Human Rights', in Davion and Wolf (2000), pp. 73-89.
- Buchanan, James (1977). *The Limit of Liberty* (Chicago: University of Chicago Press).
- (1988). 'The Gauthier Enterprise' in Paul, Miller, Paul and Ahrens (1988), pp. 75-94.
- Cohen, G. A. (2008). *Rescuing Justice and Equality* (Cambridge, MA.: Harvard University Press)
- Cohen, Joshua (2003). 'For a Democratic Society', in Freeman (2003a), pp. 86-138.
- (2010). *Rousseau: A Free Community of Equals* (New York: Oxford University Press)

Bibliography

- Coole, Diana (1994) 'Woman, Gender and Contract,' in Boucher and Kelly (1994), pp. 191-210.
- Copp, David (1991). 'Contractarianism and Moral Skepticism', in Vallentyre (1991), pp. 196-228.
- Cullity, Garrett and Berys Gaut (ed.) (1997). *Ethics and Practical Reason* (Oxford: Clarendon Press).
- Daniels, Norman (ed.) (1975). *Reading Rawls* (Oxford: Blackwell Publishing).
- Danielson, Peter (1991). 'The Lockean Provsio', in Vallentyre (1991), pp. 291-322.
- Darwall, Stephen (1976). 'A Defense of the Kantian Interpretation', *Ethics*, 86(2): 164-170
- (ed.) (2003a). *Contractarianism/ Contractualism* (Malden, MA.: Blackwell Publishing), pp. 91-107.
- (2003b). 'Introduction', in Darwall (2003a), pp. 1-8.
- (2006). 'Morality and Practical Reason: A Kantian Approach,' in David Copp (ed.), *The Oxford Handbook of Ethical Theory* (New York: Oxford University Press), P. 282-320
- (2009). *Second-Person Standpoint* (Cambridge, MA.: Harvard University Press)
- Dauenhauer, Bernard P. (2000). 'A Good Word for a Modus Vivendi', in Davion and Wolf (2000), pp. 204-220.
- Davion, Victoria, and Clark Wolf (ed.) (2000). *The Idea of a Political Liberalism* (New York: Rowman & Littlefield Publishers)
- Debreu, Gerard (1959). *Theory of Value* (New York: Wiley).
- Deigh, John (1983). 'Shame and Self-Esteem: A Critique', *Ethics*, 93(2): 225-245
- Dreben, Burton (2003). 'On Rawls and Political Liberalism', in Freeman (2003a), pp. 316-346.
- Dreier, James (ed.) (2006). *Contemporary Debates in Moral Theory* (Malden MA.: Blackwell Publishing).
- Dworkin, Ronald (1973). 'The Original Position' in Daniels 1975, pp. 16-53
- Dworkin, Gerald (2002). 'Contractualism and the Normativity of Principles', *Ethics*, 112(3): 471-482.
- Elster, Jon (2008). *Reason and Rationality*, trans. Steven Rendall (Princeton, N.J.: Princeton University Press)
- Eshete, Andreas (1974). 'Contractarianism and the Scope of Justice', *Ethics*, 85(1): 38-49.
- Finlay, Stephen (2010). 'Recent Work on Normativity,' *Analysis*, 70(2): 331-346.
- Finkelstein, Claire (2001). 'Rational Temptation', in Morris and Ripstein (2001), pp. 56-80.
- Finnis, John (1980). *Natural Law and Natural Rights* (Oxford: Clarendon Press)
- Fishkin, James (1988). 'Bargaining, Justice, and Justification: Towards Reconstruciotn' in Paul, Miller, Paul and Ahrens (1988), pp. 46-64.
- (1990). 'Towards A New Social Contract', *Nous*, 24(2): 217-226.
- (1992). *The Dialogue of Justice: Toward a self-reflective society* (New Haven and London: Yale University Press)
- Flikschuh, Katrin (2000). *Kant and Modern Political Philosophy* (Cambridge: Cambridge University Press)

Bibliography

- Foot, Philippa (2002). 'Virtues and Vices', in Philippa Foot, *Virtues and Vices* (Oxford: Clarendon Press), pp.1-18.
- Freeman, Samuel (1990). 'Reason and Agreement in Social Contract View', in Freeman (2007a), pp. 17-44.
- (1991). 'Contractualism, Moral Motivation, and Practical Reason', *The Journal of Philosophy*, 88(6); 281-303.
- (ed.) (2003a). *The Cambridge Companion to Rawls* (New York: Cambridge University Press)
- (2003b). 'Congruence and the Good of Justice', in Freeman (2003a), pp. 277-315.
- (2006). 'Moral Contractarianism as a Foundation for interpersonal Morality,' in Dreier (2006), pp. 57-76.
- (2007a). *Justice and the Social Contract* (New York: Oxford University Press).
- (2007b). 'The Burdens of Public Justification: Constructivism, Contractualism, and Publicity', *Politics, Philosophy & Economics* 6(1): 5-43.
- (2007c). *Rawls* (London: Routledge).
- Friedman, Marilyn (2000). 'John Rawls and the Political Coercion of Unreasonable People', in Davion and Wolf (2000), pp. 16-33.
- Gaus, Gerald (1990). *Value and Justification* (Cambridge, U.K.: Cambridge University Press)
- (2011). *The Order of Public Reason* (Cambridge, U. K.: Cambridge University Press)
- Gauthier, David (1974). 'Justice and Natural Endowment: Toward a Critique of Rawls's Ideological Framework,' in Gauthier (1990), pp. 150-170.
- (1975). 'Reason and Maximization', in Gauthier (1990), pp. 209-233.
- (1977). 'The Social Contract as Ideology', in Gauthier (1990), pp. 325-354.
- (1979). 'David Hume, Contractarianism', in Gauthier (1990), pp. 45-76.
- (1982). 'Three against Justice: The Foole, the Sensible Knave, and the Lydian Shepherd', in Gauthier (1990), pp. 129-149.
- (1984). 'The Incomplete Egoist', in Gauthier (1990), pp. 234-273.
- (1986). *Morals by Agreement* (Oxford: Clarendon Press).
- (1988). 'Morality, Rational Choice, and Semantic Representation: A Reply to my Critics' in Paul, Miller, Paul and Ahrens (1988), pp. 173-221.
- (1990). *Moral Dealing: Contract, Ethics, and Reason* (Ithaca: Cornell university Press)
- (1991a). 'Why Contractarianism?', in Darwall (2003a), pp. 91-107.
- (1991b). 'Rational constraint: Some last words', in Vallentyre (1991), pp. 323-330.
- and Robert Sugden (ed.) (1993). *Rationality, Justice and the Social Contract* (Hemel Hempstead : Harvester Wheatsheaf)
- (1993). 'Uniting separate persons', in Gauthier and Sugden (1993), pp. 176-192.
- (1997). 'Political Contractarianism', *Journal of Political Philosophy*, 5(2): 132-148.

Bibliography

- Gibbard, Alan (1982). 'Human Evolution and the Sense of Justice', *Midwest Studies in Philosophy*, 7(1): 31-46.
- (1985). 'Reply to Sturgeon,' *Ethics*, 96(1): 34-41.
- Glimcher, Paul W., Michael C. Dorris and Hannah M. Bayer (2005). 'Physiological utility theory and the neuroeconomics of choice,' *Games and Economic Behaviour*, 52(2): 213-256
- Goodin, Robert E. (1993). 'Equal rationality and initial endowment' in Gauthier and Sugden (1993), pp. 116-130.
- , Philip Pettit and Thomas Pogge (ed.) (2007). *A Companion to Contemporary Political Philosophy*, 2nd edition (Oxford: Blackwell Publishing).
- Gough, J. W. (1957). *The Social Contract* (Oxford: Clarendon Press).
- Greene, Joshua D., Leigh E. Nystrom. Andrew D. Engell, John M. Dorely and Jonathan D. Cohen (2001). "The Neural Bases of Cognitive Conflict and Control in Moral Judgment." *Neuron*, vol. 44 (2): 389-400
- (2003). 'From neutral "is" to moral "ought": what are the moral implications of neuroscientific moral psychology?', *Natural Review Neuroscience*, 4: 847-851.
- Griffin, James (1986). *Well-being* (Oxford: Clarendon Press)
- Grotius, Hugo (2005). *Right of War and Peace*, ed. Richard Tuck (Indianapolis: Liberty Fund).
- Hamlin, Alan (1989) 'Liberty, contract and the state,' in Alan Hamlin and Philip Pettit (ed.), *The Good Polity* (Malden, MA.: Blackwell Publishing)
- Hampton, Jean (1980). 'Contracts and Choice: Does Rawls Have a Social Contract Theory?', *Journal of Philosophy*, 77(6): 315-338.
- (1986). *Hobbes and the Social Contract Tradition* (Cambridge, U.K.: Cambridge University Press).
- (1991). 'Two faces of contractarian thought' in Vallentyne (1991), pp. 31-55.
- (1993a). 'Contract and Consent', in Goodin, Pettit and Pogge (2007), pp. 478-492.
- (1993b). 'Feminist Contractarianism' in Hampton (2007), pp. 1-38
- (1997). *Political Philosophy* (Boulder, Colo : Westview Press)
- (2007). *The Intrinsic Worth of Persons* (New York: Cambridge University Press)
- Hardin, Russell (1988). 'Bargaining for Justice' in Paul, Miller, Paul and Ahrens (1988), pp. 65-74.
- Harman, Gilbert (1977). *The Nature of Morality* (New York: Oxford University Press)
- (1988). 'Rationality in Agreement: A Commentary on Gauthier's Morals by Agreement' in Paul, Miller, Paul and Ahrens (1988), pp. 1-16.
- Harsanyi, John (1958). 'Ethics in Terms of Hypothetical Imperatives' in Harsanyi (1976), pp. 24-37.
- (1975). , 'Can the maximin Principle serve as a Basis for Morality? A Critique of John Rawls's Theory', in Harsanyi (1976), pp. 37-63.
- (1976). *Essays on ethics, social behavior, and scientific explanation* (Dordrecht, Netherland: Reidel Publishing).
- (1977). 'Rule Utilitarianism and Decision Theory', *Erkenntnis*, 11(1): 25-53.

Bibliography

- (1982). 'Morality and the Theory of Rational Behaviour,' in Sen and Williams (1982), pp. 39-62.
- (1986). *Rational Behaviour and Bargaining Equilibrium in Games and Social Situation* (Cambridge, U.K.: Cambridge University Press)
- Hart, H. L. A. (1982). *Essays on Bentham* (Oxford: Clarendon Press)
- Hegel, Georg Wilhelm Friedrich (1991). *Elements of the Philosophy of Right* (Cambridge, U.K.: Cambridge University Press)
- Held, Virginia (1976). 'On Rawls and Self-interest', *Midwest Studies in Philosophy*, 1(1): 57-60.
- (1977). 'Rationality and Reasonable Cooperation', *Social Research*, 44 (4): 708-744
- Hershovitz, Scott (2000). 'A Mere Modus Vivendi?', in Davion and Wolf (2000), pp. 221-230.
- Hill, Thomas (1994). 'The Stability Problem in *Political Liberalism*', *Pacific Philosophical Quarterly*, 75: 333-352.
- Hobbes, Thomas (1994). *Leviathan*, ed. Edwin Curley (Cambridge, U. K.: Hackett Publishing).
- Hooker, Brad (2000). *Ideal Code, Real World* (New York: Oxford University Press).
- (2002). 'Contractualism, Spare Wheel, Aggregation', in Matravers (2003), pp. 53-76.
- Horton, John (1992). *Political Obligation* (Basingstoke: Macmillan)
- Hubin, Donald, and Mark Lambeth (1991). 'Providing for Rights', in Vallentyre (1991), pp. 112-126.
- Hume, David (1994). 'Of the Original Contract,' in David Hume, *Political Essays*, ed. Knud Haakonssen (Cambridge, U.K.: Cambridge University Press), pp. 186-201.
- (2000). *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton (Oxford: Oxford University Press)
- James, Aaron (2004). 'Rights and Circularity in Scanlon's Contractualism,' *Journal of Moral Philosophy*, 1(3): 369-377.
- Kahn, Charles (1981). 'The Origins of Social Contract Theory,' in G.B. Kerferd (ed.), *The Sophists and Their Legacy* (Wiesbaden, Germany: Franz Steiner Verlag), pp. 92-108.
- Kant, Immanuel (1991a). 'On The Common Saying: "This May Be True In Theory, But It Does Not True In Practice"', in Immanuel Kant, *Political Writings*, trans. H. S. Reiss, 2nd edition (Cambridge, UK.: Cambridge University Press), pp. 61-92.
- (1991b). 'Perpetual Peace: A Philosophical Sketch', in Immanuel Kant, *Political Writings*, trans. H. S. Reiss, 2nd edition (Cambridge, UK.: Cambridge University Press), pp. 93-130.
- (1998). *Groundwork of the Metaphysics of Morals*, trans. Mary Gregor (Cambridge, UK.: Cambridge University Press).
- Kavka, Gregory (1986). *Hobbesian Moral and Political Theory* (Princeton, N. J.: Princeton University Press).

Bibliography

- Kelly, Paul (1994). 'Justifying "Justice": Contractarianism, Communitarianism and the Foundation of Contemporary Liberalism', in Boucher and Kelly (1994), pp. 226-244.
- Klosko, George (1994). 'Rawls's Argument from Political Stability', *Columbia Law Review*, 94: 1882-1897.
- (2000). *Democratic Procedures and Liberal Consensus* (Oxford: Oxford University Press)
- (2009). 'Rawls's Public Reason and American Society', in Shaun Young (ed.), *Reflections on Rawls* (Farnham; Burlington: Ashgate Publishing).
- Korsgaard, Christine (1996). *The Source of Normativity*, ed. Onora O'Neill (Cambridge, U.K.: Cambridge University Press)
- (1999). 'The Myth of Egoism', in Korsgaard (2008), pp. 69-99.
- (2008). *The Constitution of Agency* (New York: Oxford University Press)
- Kraus, Jody S and Jules Coleman (1991). 'Morality and the Theory of Rational Choice', in Vallentyre (1991), pp. 254-290.
- (1993). *The Limit of Hobbesian Contractarianism* (Cambridge, U.K.: Cambridge University Press)
- Kukathas, Chandran (2001). 'Education and citizenship in diverse societies', *International Journal of Educational Research*, 35: 319-330.
- (ed.) (2003a). *John Rawls: Critical assessments of leading political philosophers, Vol. 1* (London: Routledge)
- — (2003b). *John Rawls: Critical assessments of leading political philosophers, Vol. 3* (London: Routledge)
- — (2003c). *John Rawls: Critical assessments of leading political philosophers, Vol. 4* (London: Routledge)
- Kumar, Rahul (1999). 'Defending the Moral Moderate: Contractualism and Common Sense', *Philosophy and Public Affairs*, 28(4): 275-309.
- Kymlicka, Will (1993). 'The Social Contract Tradition', in Peter Singer (ed.), *A Companion to Ethics* (Oxford: Blackwell Publishing), pp.186-196.
- Larmore, Charles (2003). 'Public Reason' in Freeman (2003a), pp. 368-393.
- Lehning, Percy (1993). 'Right constraints?', in Gauthier and Sugden (1993), pp. 95-115.
- Lessnoff, Michael (1986). *Social Contract* (London: Macmillan).
- Locke, John (1988). *Two Treatises of Government*, ed. Peter Laslett, 3rd edition (Cambridge, UK.: Cambridge University Press).
- Loomes, Graham and Robert Sugden (1982). 'An Alternative Theory of Rational Choice under Uncertainty', *The Economic Journal*, 92(368): 805-824.
- Luce, Duncan, and Raiffa Howard (1957). *Games and Decisions: An Introduction and Critical Survey* (New York: Wiley)
- Ludwig, Kirk (2004). 'Rationality, Language, and the Principle of Charity,' in Alfred Mele and Piers Rawling (ed.), *The Oxford Handbook of Rationality* (New York: Oxford University Press), pp. 343-362.
- Macedo, Stephen (2000). *Diversity and Distrust* (Cambridge, MA.: Harvard University Press).
- MacIntyre, Alasdair (1985). *After Virtue*, 2nd edition (London: Duckworth)

Bibliography

- Mackie, J. L. (1990). *Ethics: Inventing Right and Wrong* (Harmondsworth: Penguin).
- Marshall, John (1975). "The Failure of Contract as Justification". *Social theory and practice*, 3 (4): 441-459.
- Marshall, Sarah (2003). 'Scanlon and Reason' in Matravers (2003), pp. 13-32.
- Marx, Karl (2000). 'On the Jewish Question,' in David McLellan (ed.), *Karl Marx: Selected Writings*, 2nd edition (Oxford: Oxford University Press), pp. 21-53.
- Matravers, Matt (ed.) (2003). *Scanlon and Contractualism* (London: Frank Cass).
- McClennen, Edward (1988). 'Constrained Maximization and Resolute Choice', in Miller, Paul and Ahrens (1988), pp. 95-11
- (2001). 'The Strategy of Cooperation', in Morris and Ripstein (2001), pp. 189-208.
- Mendus, Susan (2003). 'The magic in the Pronoun "My",' in Matravers (2003), pp. 33-52.
- Miller, Dale (1999). 'The Norm of Self-interest,' *American Psychologists*, 54 (12): 1053-1060
- Miller, David (ed.) (1987). *The Blackwell Encyclopedia of Political Thought* (Oxford: Blackwell Publishing)
- Mills, Claudia (2000). "'Not a Mere Modus Vivendi": The Bases for Allegiance to the Just State', in Davion and Wolf (2000), pp. 190-203.
- Milo, Ronald (1995). 'Contractarian Constructivism', *The Journal of Philosophy*, 92(4): 181-204.
- Morris, Christopher W. (1988). 'The Relation between Self-interest and Justice in Contractarian Ethics,' *Social Philosophy and Policy*, 5(2): 119-153.
- (1991). 'Moral standing and rational-choice contractarianism', in Vallentyre (1991), pp. 76-95.
- and Arthur Ripstein (2001). *Practical Rationality and Preference: Essays for Gauthier* (New York: Cambridge University Press)
- Morton, Adam (2001). 'Psychology for Cooperators', in Morris and Ripstein (2001), pp. 153-172.
- Mulhall, Stephen, and Adam Swift (1996). *Liberals and Communitarians* (Oxford: Blackwell Publishing)
- Nagel, Thomas (1970). *The Possibility of Altruism* (Oxford: Clarendon Press)
- (1973). 'Rawls on Justice', in Daniel (1975), pp. 1-16.
- (1986). *The View from Nowhere* (New York: Oxford University Press)
- (1991). *Equality and Partiality* (New York: Oxford University Press).
- (2003). 'Rawls and Liberalism', in Freeman (2003a), pp. 62-85.
- Narveson, Jan (1991). 'Gauthier on Distributive Justice and the Natural Baseline', in Vallentyre (1991), pp. 127-147.
- Nida-Rumelin, Julian (1991). 'Practical reason, collective rationality and contractarianism', in Gauthier and Sugden (1993), pp. 53-74.
- Nozick, Robert (1974). *Anarchy, State, and Utopia* (Oxford: Basil Blackwell).
- Nussbaum, Martha (2006). *The Frontier of Justice* (Cambridge, MA.: The Belknap Press of Harvard University Press).
- Oakeshott, Michael (1975). *Hobbes on Civil Association* (Oxford: Blackwell Publishing)

Bibliography

- Okin, Susan Moller (1989). *Justice, Gender, and the Family* (Basic Books)
- O'Neill, Onora (2004). 'Constructivism VS. Contractarianism,' in Stratton-Lake (2004), pp. 19-31.
- Paul, Ellen Frankel, Fred D. Miller Jr., Jeffrey Paul, John Ahrens (ed.) (1988). *The New Social Contract: Essays on Gauthier* (Oxford: Basil Blackwell)
- Pateman, Carole (1985). *The Problem of Political Obligation* (Cambridge, U. K.: Polity)
- (1988). *The Sexual Contract* (Cambridge, U.K.: Polity)
- and Charles Mills (2007). *Contract and Domination* (Cambridge, U.K.: Polity)
- Peters, R.S. (1956). *Hobbes* (Harmondsworth: Penguin).
- Pettit, Philip (2006). 'Can Contract Theory ground morality?', in Dreier (2006), pp. 77-96.
- Pitkin, Hannah (1972). 'Obligation and Consent', in Peter Laslett, W.G. Runciman, and Quentin Skinner (eds.), *Philosophy, Politics and Society*, 4th series (Oxford: Blackwell Publishing), pp. 45-85.
- Pogge, Thomas (1989). *Realizing Rawls* (London: Cornell University Press).
- (2001). 'What we can reasonably reject?', *Nous*, 35(s1): 118-147.
- Pufendorf, Samuel (1991). *On the Duty of Man and Citizens According to Natural law*, trans. Michael Silverthorne (Cambridge, UK.: Cambridge University Press).
- Quong, Jonathan (2011). *Liberalism Without Perfection* (Oxford: Oxford University Press).
- Railton, Peter (1997). 'On the Hypothetical and Non-hypothetical in Reasoning about Belief and Action,' in Cullity and Gaut (1997), pp. 53-80.
- Rapoport Anatol, and Albert Chammah (1965). *Prisoner's Dilemma: A Study in Conflict and Cooperation* (Ann Arbor: University of Michigan Press)
- Rawls, John (1958). 'Justice as Fairness' in Rawls (1999b), pp. 47-72.
- (1963). 'The Sense of Justice', in Rawls (1999b), pp. 47-72.
- (1972). *A Theory of Justice* (Oxford: Clarendon Press).
- (1975). 'Fairness to Goodness', in Rawls (1999b), pp. 167-185.
- (1980). 'Kantian Constructivism in Moral Theory' in Rawls (1999b), pp. 303-358.
- (1985). 'Justice as Fairness: Political not Metaphysical' in Rawls (1999b), pp. 388-414.
- (1989a). 'Themes in Kant's Moral Philosophy', in Rawls (1999b), pp. 497-528.
- (1989b). 'The Domain of the Political and Overlapping Consensus', in Rawls (1999b), pp. 473-496.
- (1993). *Political Liberalism* (New York: Columbia University Press).
- (1997). 'The Idea of Public Reason Revisited' in Rawls (1999b), pp. 573-615.
- (1999a). *A Theory of Justice*, revised edition (Cambridge, MA.: The Belknap Press of Harvard University Press).
- (1999b). *Collected Papers*, ed. Samuel Freeman (Cambridge, MA.: The Belknap Press of Harvard University Press)
- (1999c). *The Law of People* (Cambridge, MA.: Harvard University Press)
- (2001). *Justice as Fairness: a Restatement* (Cambridge, MA : The Belknap Press of Harvard University Press)

Bibliography

- (2005). *Political Liberalism*, expanded edition (New York: Columbia University Press)
- (2007). *Lectures on the History of Political Philosophy*, ed. Samuel Freeman (Cambridge, MA.: The Belknap Press of Harvard University Press).
- Raz, Joseph (1979a). *The Authority of Law* (Oxford: Clarendon Press).
- (1979b). 'Reasons for Action, Decision and Norms,' in Joseph Raz (ed.), *Practical Reasoning* (Oxford: Oxford University Press), pp. 128-143.
- (1990). 'Facing diversity: the case of epistemic abstinence', in Kukathas (2003c), pp. 395-431.
- Richards, D. A. J. (1971). *A Theory of Reason for Action* (Oxford: Clarendon Press).
- Richardson, Henry S., and Paul Weithman (ed.) (1999a). *The Philosophy of Rawls: A Collection of Essays, Vol. 1* (New York: Garland)
- (1999b). *The Philosophy of Rawls: A Collection of Essays, Vol. 2* (New York: Garland)
- (1999c). *The Philosophy of Rawls: A Collection of Essays, Vol. 4* (New York: Garland)
- Riley, Patrick (1982). *Will and Political Legitimacy* (Cambridge, MA.: Harvard University Press).
- Ripstein, Arthur (2001). 'Preference', in Morris and Ripstein (2001), pp. 37-55.
- Rousseau, Jean-Jacques (1979). *Emile*, trans. Allan Bloom (Basic Books).
- (1987). 'The Social Contract' in Jean-Jacques Rousseau, *The Basic Political Writings*, trans. Donald Cress (Indianapolis: Hackett Publishing).
- Ryan, Alan (1988) 'Hobbes and Individualism,' in G. A. J. Rogers and Alan Ryan (ed.), *Perspective on Thomas Hobbes* (Oxford: Clarendon Press)
- Sandel, Michael (1982). *Liberalism and the Limit of Justice* (Cambridge, U.K.: Cambridge university Press)
- (2006). *Public Philosophy: Essays on Morality in Politics* (Cambridge, MA.: Harvard University Press)
- Sayre-McCord, Geoffrey (1991). 'Deception and reasons to be moral' in Vallentyre (1991), pp. 181-195.
- (2000). 'Contractarianism', in Hugh LaFollette (ed.), *The Blackwell Guide to Ethical Theory* (Malden, MA.: Blackwell Publishing), pp. 247-267
- Scheffler, Samuel (1978). 'Moral Independence and the Original Position', *Philosophical Studies*, 35(4): 397-403.
- (1992). *Human Morality* (New York: Oxford University Press)
- Scanlon, T. M. (1972). 'A theory of freedom of expression' in Scanlon (2003a), pp. 6-25.
- (1975). 'Preference and urgency', in Scanlon (2003a), pp. 70-83.
- (1977). 'Due Process', in Scanlon (2003a), pp. 42-69.
- (1978). 'Rights, goals, and fairness' in Scanlon (2003a), pp. 26-41.
- (1982). 'Contractualism and Utilitarianism', in Sen and Williams (1982), pp. 103-128.
- (1988). 'The Significance of Choice,' in *The Tanner Lectures on Human Values* (Salt Lake City, UT.: University of Utah Press, 1988)
- (1993). 'Value, desire, and quality of life', in Scanlon (2003a), pp. 169-186.

Bibliography

- (1996). 'The difficulty of tolerance' in Scanlon (2003a), pp. 187-201.
- (1998). *What We Owe to Each Other* (Cambridge, MA.: The Belknap Press of Harvard University Press).
- (2000). 'A Contractualist Reply', *Theoria*, 66(3): 237-245.
- (2001). 'Promises and Contracts', in Scanlon (2003a), pp. 234-269.
- (2003a). *The Difficulty of Tolerance* (Cambridge, U.K.: Cambridge University Press)
- (2003b). 'Reply to Gauthier and Gibbard,' *Philosophy and Phenomenological Research*, 66(1): 176-189.
- (2003c). 'Rawls on Justification', in Freeman (2003a), pp. 139-167.
- (2004). 'Replies' in Stratton-Lake (2004), pp. 123-138.
- (2008). *Moral Dimensions* (Cambridge, MA.: The Belknap Press of Harvard University Press).
- (2009). 'The Kingdom of Ends on the cheap,' in Alex Voorhoeve, *Conversations on Ethics* (New York: Oxford University Press)
- Sen, Amartya and Bernard Williams (ed.) (1982). *Utilitarianism and Beyond* (Cambridge, U.K.: Cambridge University Press)
- (1977). 'Rational Fool', *Philosophy & Public Affairs*, 6(4): 317-344
- Shackle, George (1964). 'General Thought-schemes and the Economist', *Woolwich Economic Paper 2* (Woolwich, NJ.: Woolwich Polytechnic Press)
- Shklar, Judith (1969). *Men and Citizens: A Study of Rousseau's Social Theory* (London: Cambridge University Press)
- Sibley, W. M. (1953). 'The Rational Versus the Reasonable', *The Philosophical Review*, 62(4): 554-560
- Sidgwick, Henry (1962). *The Method of Ethics*, 7th edition (London: Macmillan).
- Simmons, A. J. *Moral Principles and Political Obligations* (Princeton, N.J.: Princeton University Press)
- (1993a). *On the Edge of Anarchy* (Princeton, N.J.: Princeton University Press)
- (1993b). 'Liberal Impartiality and Political Legitimacy', *Philosophical Books*, 34(4): 213-223.
- Simon, Herbert (1962). *The Science of the Artificial* (Cambridge, MA.: MIT Press).
- Skyrms, Brian (1996). *Evolution of the Social Contract* (Cambridge, U.K.: Cambridge University Press)
- Southwood, Nicholas (2010). *Contractualism and the Foundation of Morality* (Oxford: Oxford University Press).
- Sugden, Robert (1993). 'Rationality and Impartiality: Is the contractarian enterprise possible?', in Gauthier and Sugden (1993), pp. 157-175.
- Starks, Cynthia A. (2000). 'Hypothetical Consent and Justification,' *The Journal of Philosophy*, 97(6): 313-334
- (2007). 'How to Include the Severely Disabled in a Contractarian Theory of Justice', *The Journal of Political Philosophy*, 15(2): 127-45.
- Stratton-Lake, Philip (ed.) (2004). *On What We Owe to Each Other* (Oxford: Blackwell Publishing)
- Taylor, Charles (1975). *Hegel* (Cambridge, U.K.: Cambridge University Press)

Bibliography

- Thompson, Judith Jarvis (1990) *The Realm of Rights* (London: Harvard University Press)
- Vallentyne, Peter (1991a). *Contractarianism and Rational Choice: Essays on Gauthier's Morals by Agreement* (Cambridge, U.K.: Cambridge University Press).
- (1991b). 'Contractarianism and the assumption of mutual unconcern', in Vallentyne (1991a), pp. 71-75.
- (1991c). 'Gauthier's Three Projects', Vallentyne (1991a), pp. 1-11.
- Wallace, Jay (2002). 'Scanlon's Contractualism' *Ethics*, 112(3): 429-470.
- Watson, Gary (1998). 'Some Considerations in Favor of Contractualism,' in Darwall (2003a), pp. 249-269.
- Weale, Albert (1993). 'Justice, Social Union and the separateness of persons', in Gauthier and Sugden (1993), pp. 75-94.
- Weithman, Paul (2010). *Why Political Liberalism?: On John Rawls's Political Turn* (New York: Oxford University Press)
- Wenar, Leif (1995). 'Political Liberalism: an internal critique', in Kukathas (2003c), pp. 57-85.
- Wolf, Clark (2000). 'Fundamental Rights, Reasonable Pluralism, and the Moral Commitments of Liberalism', in Davion and Wolf (2000), pp. 102-126.
- Wolff, Robert Paul (1977). *Understanding Rawls* (Princeton, N.J.; Princeton University Press)
- (1998). *In Defence of Anarchism* (Berkeley and Los Angeles, California: University of California Press)
- Wolff, Jonathan (2004). 'Scanlon on Well-Being', in Stratton-Lake (2004), pp. 32-45.
- (2006). *An Introduction to Political Philosophy*, revised edition (Oxford: Oxford University Press)
- Wright, G. H. von (1983). 'On so-called practical inference', in G. H. von Wright, *Philosophical papers Vol.1: Practical reason* (Oxford: Blackwell Publishing), pp. 56-75